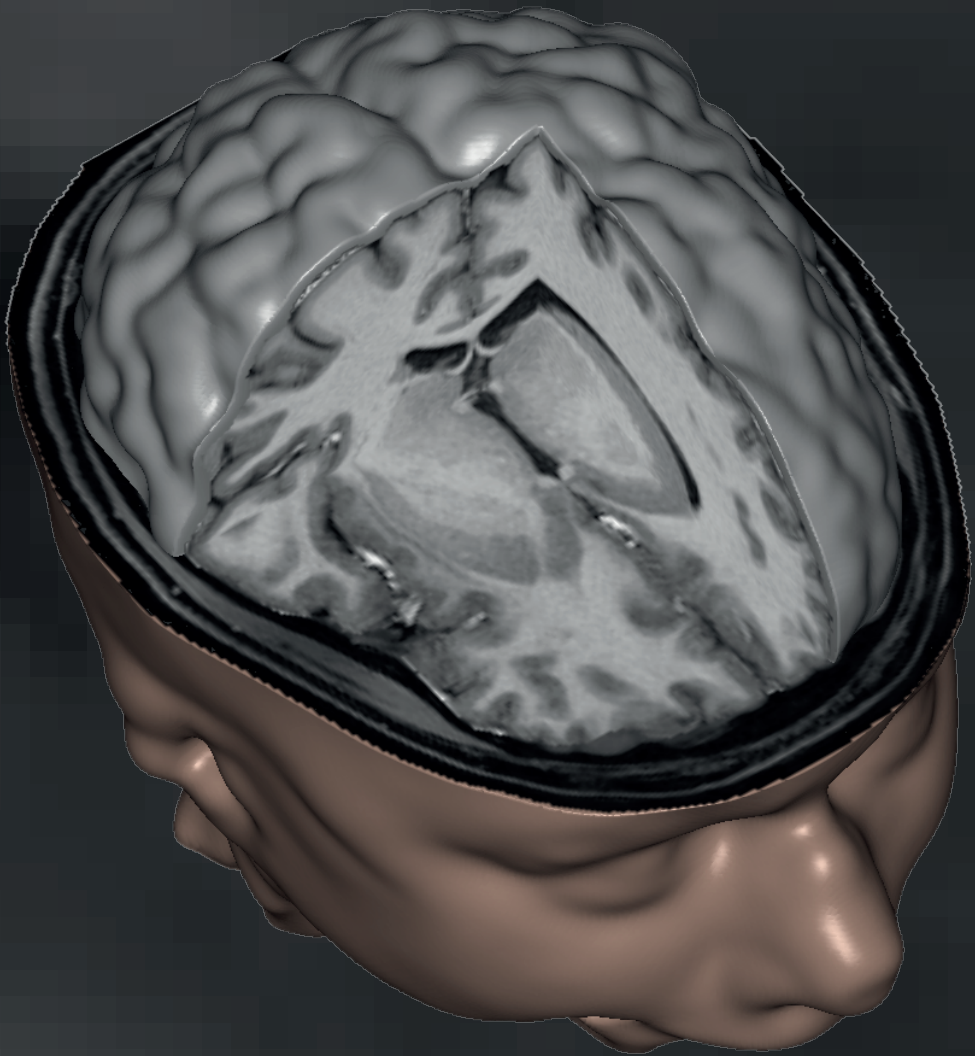# Towards earlier detection of Alzheimer's disease using Magnetic Resonance images

Rita Simões

TOWARDS EARLIER DETECTION OF ALZHEIMER'S DISEASE USING MAGNETIC RESONANCE IMAGES

Rita Simões

Graduation committee:

Chairman:	Prof. dr. ir. A.J. Mouthaan
Promoter:	Prof. dr. ir. C.H. Slump
Assistant promoter:	Dr. A-M. van Cappellen van Walsum

Members:
Prof. dr. C.F. Beckmann	University of Twente
Dr. med. C. Mönninghoff	Essen University Hospital
Prof. dr. ir. A. Pižurica	Ghent University
Prof. dr. ir. P.H. Veltink	University of Twente
Prof. dr. ir. P.H.N. de With	Eindhoven University of Technology

TOWARDS EARLIER DETECTION OF ALZHEIMER'S DISEASE USING MAGNETIC RESONANCE IMAGES

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the Rector Magnificus,
Prof. dr. H. Brinksma,
on account of the decision of the graduation committee,
to be publicly defended
on Thursday 21 November, 2013 at 12:45

by

Ana Rita Lopes Simões

born on 29 May, 1987
in Viseu, Portugal

This dissertation has been approved by:

Prof. dr. ir. C.H. Slump (promoter)
Dr. A.-M. van Cappellen van Walsum (assistant promoter)

Abstract

Alzheimer's disease (AD) is the most common type of dementia and a major cause of disability worldwide. Early detection of AD is essential to provide the patients with adequate and timely treatments and to help researchers monitor their effectiveness. Structural Magnetic Resonance Imaging (MRI) is a diagnostic tool that provides high-resolution images and a high brain tissue contrast.

MRI-based biomarkers have been investigated in an attempt to describe and quantify structural differences between groups of normal elderly controls and subjects suffering from AD. Additionally, classification methods have been proposed that use these biomarkers as features to distinguish between those groups, thereby also providing diagnostic value.

Two main approaches have been extensively explored in the past decades to perform early-stage AD classification based on structural MR images. The first uses the volume and/or the shape of specific brain structures, such as the hippocampi and the entorhinal cortex. As a consequence, these methods rely substantially on the quality of: 1) the assumptions of which brain regions are affected at an early stage of AD; 2) the segmentation of these brain structures, which suffers from large variability across studies. Another major line of research overcomes the first drawback by using voxelwise measures, such as the probability maps of the brain tissues. However, these methods require a voxelwise inter-subject correspondence, which is difficult to achieve, particularly considering the large anatomical variability of the brain across different subjects.

Besides the above-mentioned disadvantages of these two approaches, they both focus on structural (volume, shape, density) changes only. It has recently been considered that also the MR image intensities and textures can provide complementary information that is overlooked by the structural-based features.

In this thesis, we propose methods to help diagnose AD at an early stage of development. In particular, we build on the existing literature on classification approaches that use MR image textures for early detection of AD.

Firstly, we focus our analysis on a type of lesions in the white matter (white matter hyperintensities) that have been shown to play a role in cognitive decline. We propose a method to automatically segment these lesions from a single MRI modality that can be suitable for large-scale clinical trials. We show that our method, despite using less information, performs similarly to current state-of-the-art multimodal approaches. Afterwards, we evaluate the performance of white matter lesion texture descriptors in the detection of Mild Cognitive Impairment (MCI, a transitional stage between normal ageing and dementia). Results show that the textures are more discriminative than the widely used lesion volumes and locations.

Secondly, we evaluate three approaches that use texture descriptors without requiring prior brain structure segmentations. The first one takes the gray-level histograms computed in the whole brain and in several cubic image regions (patches). The second approach considers second-order statistical texture maps and the third one uses intensity-invariant texture descriptors. Similarly to the first method, these are also determined at cubic local patches.

The results from these three approaches show that: 1) texture descriptors are able to achieve high classification rates, comparably to (or better than) structural-based features; 2) by using local patches over the entire brain, no assumptions need to be made about the expectedly affected brain regions, and consequently no prior segmentations are needed; 3) by only affine-registering the images (without performing non-linear alignments) we are still able to localize discriminative brain regions using finely sampled patches in the brain.

Samenvatting

De ziekte van Alzheimer (AD) is de meest voorkomende vorm van dementie en een belangrijke oorzaak van invaliditeit in de wereld. Vroege detectie van AD is essentieel om de patiënten te voorzien van adequate en tijdige behandeling en om onderzoekers te helpen hun effectiviteit te bewaken. Structurele Magnetic Resonance Imaging (MRI) is een diagnostisch instrument dat beelden biedt met hoge resolutie en een hoog hersenweefsel contrast.

MRI-gebaseerde biomarkers zijn onderzocht in een poging om structurele verschillen te beschrijven en te kwantificeren tussen groepen normale controles en patiënten die lijden aan AD. Daarnaast zijn classificatie methoden voorgesteld die gebruik maken van deze biomarkers als kenmerken om een onderscheid te maken tussen die groepen, waardoor ook het verstrekken van diagnostische waarde.

Twee belangrijke benaderingen zijn uitgebreid onderzocht in de afgelopen decennia om vroeg AD-classificatie op basis van structurele MR beelden uit te voeren. De eerste maakt gebruik van het volume en/of de vorm van specifieke hersenstructuren zoals de hippocampus en de entorhinale cortex. Bijgevolg zijn deze methoden sterk afhankelijk van de kwaliteit van: 1) de aannames over welke hersengebieden getroffen zijn in een vroeg stadium van AD, 2) de segmentering van deze hersenstructuren, die lijdt aan grote variatie tussen studies. Een andere belangrijke lijn van onderzoek overwint de eerste nadeel met voxelgewijze maatregelen, zoals de waarschijnlijkheid kaarten van de hersenweefsels. Deze werkwijzen vereisen een voxelgewijze interindividuele correspondentie, die moeilijk te bereiken is, vooral gezien het grote anatomische variabiliteit van de hersenen van verschillende proefpersonen.

Naast de bovengenoemde nadelen van deze twee benaderingen, zijn beide gericht op structurele (volume, vorm, dichtheid) wijzigingen alleen. Het is onlangs aangetoond dat ook de MR afbeelding intensiteiten en texturen aanvullende informatie kunnen bieden, die over het hoofd wordt gezien door de structurele gebaseerde kenmerken.

In dit proefschrift stellen we methoden voor het diagnosticeren van AD

in een vroeg stadium van ontwikkeling. In het bijzonder bouwen we voort op de bestaande literatuur over classificatie benaderingen die MR afbeelding texturen gebruiken voor de vroegtijdige detectie van AD.

Ten eerste richten we onze analyse op een soort laesies in de witte stof (witte stof hyperintensiteiten) die worden verondersteld om een rol te spelen in de cognitieve achteruitgang. Wij stellen een methode voor het automatisch segmenteren van deze laesies uit een MRI modaliteit die geschikt kan zijn voor grootschalige klinische trials. We zien dat onze methode, ondanks het gebruik van minder informatie, vergelijkbaar presteert aan de huidige state-of-the-art multimodale benaderingen. Daarna evalueren we de prestaties van de witte stof laesie textuur descriptoren in de opsporing van Mild Cognitive Impairment (MCI, een overgangsfase tussen normale veroudering en dementie). Resultaten tonen aan dat de textures meer onderscheidend zijn dan de veel gebruikte volumes en locaties van de laesies.

Ten tweede evalueren we drie benaderingen die textuur descriptoren gebruiken zonder voorafgaande hersenstructuur segmentaties. De eerste neemt de grijze-niveau histogrammen in het hele brein en in verschillende kubieke image regio ("patches"). De tweede benadering beschouwt tweede-orde statistische textuur kaarten en de derde maakt gebruik van intensiteit-invariant textuur descriptoren. Net als de eerste methode, worden deze ook in kubische lokale "patches" berekend.

Uit de resultaten van deze drie benaderingen blijkt dat: 1) textuur descriptoren zijn in staat om een hoge classificatie nauwkeurigheid te bereiken, vergelijkbaar met (of beter dan) de structurele gebaseerd kenmerken; 2) door het gebruik van lokale "patches" over de gehele hersenen, hebben we geen aannames nodig over de verwacht getroffen gebieden van de hersenen, en dus ook geen voorafgaande segmentaties; 3) door alleen affine-registratie van de beelden (zonder het uitvoeren van niet-lineaire registratie) zijn we nog steeds in staat om onderscheidende gebieden van de hersenen te lokaliseren met behulp van fijn bemonsterd "patches" in de hersenen.

Contents

CHAPTER 1

Introduction

1.1 Alzheimer's Disease

Alzheimer's disease (AD) is a neurodegenerative disease and the most common cause of dementia worldwide. The current prevalence of AD is about 1-2% at 65 years old and 35% or higher by age 85 [1]. As life expectancy increases, the number of people suffering from AD will grow rapidly. In 2006, the estimated number of people with AD was 26.6 million. This number is expected to quadruple by 2050, meaning that, by that time, 1 in 85 persons worldwide will suffer from AD [2]. Therefore, besides causing a major psychological burden on patients, families and caregivers, AD is also expected to place an increasingly large socioeconomic burden in our societies [3].

Clinically, AD is characterized by a gradual cognitive decline that usually starts with memory impairment (short-term memory in earlier stages) and progresses to the deterioration of functional abilities, to behavioral changes, ultimately leading to a complete loss of independence in late-stage patients [4].

The exact pathogenesis of AD is not yet fully understood, with multiple processes currently thought to be involved in the disease development. Since the early 1990's, the so-called "amyloid cascade hypothesis" has had a prominent role in describing the etiology and pathogenesis of AD. According to this hypothesis, AD starts with the accumulation of Aβ proteins in the brain. These trigger the formation of senile plaques (SP) and neurofibrillary tangles (NFT), which in turn progressively lead to damage and loss of the neural tissue and consequently to dementia [5] (Figure 1.1).

However, recent evidence shows that SP and NFT may develop independently and that they may be the result of neurodegeneration rather than its cause [6]. Other disease mechanisms have then recently been investigated in an attempt to better describe the AD pathology [1].
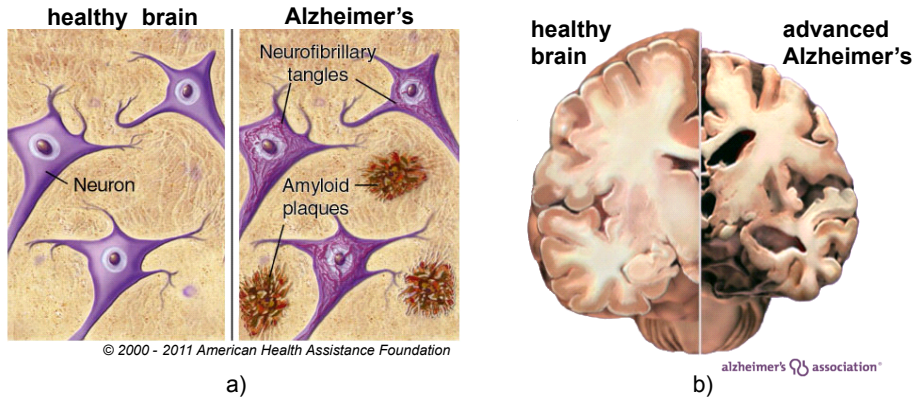
Figure 1.1: AD hallmarks: a) tissue-level representation, showing the presence of amyloid (senile) plaques and neurofibrillary tangles; b) late-stage AD brain, showing marked shrinkage in comparison with a healthy brain.

Currently, a definite diagnosis of AD will not be available until an autopsy is made, i.e., post-mortem, or, in rare cases, through a brain biopsy. These tests are able to confirm the presence of SP and NFT and consequently determine the cause of dementia as being AD [7]. Furthermore, the clinical diagnosis for "probable AD" cannot be given until the patient shows severe cognitive deficits that significantly impact his/her daily life activities [8].

However, evidence shows that AD pathology starts decades before the first symptoms arise [9]. Therefore, there is increasing interest in finding indicators ("biomarkers") of AD that can help diagnose the disease at an incipient stage. In particular, existing pharmacological therapies are only symptomatic treatments that are prescribed for later stages of AD [3]. These therapies provide temporary and modest improvement in cognitive functions but do not cure the disease [1]. An earlier diagnosis is expected to help with the proper screening of patients for clinical trials and consequently lead to the development of more suitable treatments. Additionally, an earlier intervention is likely to be more effective since it can be applied before irreversible damage has taken place [10]. It is estimated that interventions capable of delaying disease onset and progression by only one year would be able to reduce the number of AD patients in 2050 by 9.2 million worldwide [2].

Recent clinical and research guidelines for diagnosing AD consider the disease progression as consisting of three stages: a pre-clinical (pre-symptomatic)

stage; a symptomatic, pre-dementia stage called "Mild Cognitive Impairment (MCI) due to AD" and the AD or dementia stage [11].

MCI is generally considered to be a transitional stage between normal ageing and dementia, characterized by memory impairment as the most prominent feature. Because MCI subjects have been considered to be at an increased risk of developing AD, much effort has been put into distinguishing MCI individuals that will convert to AD from those that will not [12].

However, even though there are general guidelines for the diagnosis of MCI [13], some criteria are not objective, which leads to a large variability in the definition of MCI subjects across studies [14]. Also, a recent study by Morris *et al.* [15] shows that MCI subjects progress gradually to more severe stages of dementia at rates that depend on the level of cognitive impairment at baseline, suggesting that the "MCI due to AD" stage [11] represents, in reality, the earliest symptomatic stage of AD.

Despite its large heterogeneity and the controversy regarding its exact definition, MCI remains a group of interest in the study of early-stage AD.

1.2   Magnetic Resonance Imaging

Neuroimaging techniques enable *in vivo* assessment of brain changes and are therefore promising in the field of early detection of AD [16]. Earlier clinical guidelines supported the use of neuroimaging in the diagnosis of AD, mostly to rule out other (possibly treatable) causes of memory loss [17]. Nowadays, the revised criteria for AD further recommends the use of neuroimaging biomarkers in research settings to complement clinical assessments [13].

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique with widespread use in research and clinical practice. It is based on the principle of Nuclear Magnetic Resonance (NMR), in which nuclei, in the presence of an external magnetic field, absorb and re-emit electromagnetic radiation at a specific resonance frequency [18].

The human body is composed of large amounts of water molecules, which in turn contain two hydrogen protons (^{1}H) each. Nuclei with an odd number of protons and/or neutrons, such as that of hydrogen, exhibit a magnetic moment and are therefore NMR-active. When a strong external magnetic field is applied, the protons will align with the field. This alignment can be either parallel or anti-parallel to the field. The parallel alignment corresponds to a lower energy state and will therefore be more occupied than the correspond-

ing anti-parallel state, resulting in a net magnetization vector that is parallel to the magnetic field.

The magnetic moment of each active nucleus precesses around its axis at the so-called Larmor frequency. To obtain the nuclear resonance effect, a radio-frequency (RF) electromagnetic pulse with the same frequency is applied to perturb the equilibrium state of the magnetic moments. When this RF pulse is turned off, the magnetic moments will return to their equilibrium state (aligned with the strong external field) by emitting an RF signal.

In particular, a 90° pulse will orient the magnetization vector perpendicularly to the static magnetic field. The return of the longitudinal magnetization (component of the magnetization vector along the direction of the static field) to the equilibrium state, after the pulse is turned off, is referred to as longitudinal relaxation, and its time constant is called T1. At the same time, the transverse magnetization (component of the magnetization vector perpendicular to the direction of the static field), which is created when the magnetic moments are flipped by the RF pulse, will decay as the magnetic moments get out of synchronization. This decay is exponential, characterized by the time constant T2 [18].

Also, spatial information can be extracted (to build an image) by applying a controlled spatial and time-variant magnetic field, which selectively excites nuclei at specific positions in the body. The combination of the gradient fields and the applied pulses is named pulse sequence. The measured signal, which is read by an RF detector system, represents the sum of the signals emitted by active nuclei from a certain part of the tissue, selected according to the pulse sequence [18].

By varying the pulse sequences, it is possible to measure different properties of the tissues being imaged. For example, a T1-weighted image shows differences in the T1 relaxation times of the different tissues. In the particular case of brain images, T1-weighting provides good contrast between gray and white matter and is therefore widely used for brain segmentation and consequently for the assessment of brain atrophy. Similarly, T2-weighted images reflect differences in the T2 relaxation time of the tissues. This modality is able to differentiate water from fat and is therefore suitable for imaging edema. T2-weighted images have also shown to be more sensitive to microscopic neurodegenerative processes than T1 images [19].

Another MRI modality that is often used for brain imaging is Fluid-Attenuation Inversion Recovery (FLAIR). It is based on T2-weighting, with the difference that the cerebrospinal fluid signal is attenuated. This causes lesions

present in the white matter to show with increased contrast with respect to healthy tissues [20].

T1, T2 and FLAIR images are often classified as structural MRI modalities [21], since they are able to provide information about large-scale properties such as the size, shape and volume of the imaged tissues. Figure 1.2 shows examples of brain MR images (coronal slices) obtained from the same subject using T1, T2 and FLAIR pulse sequences.
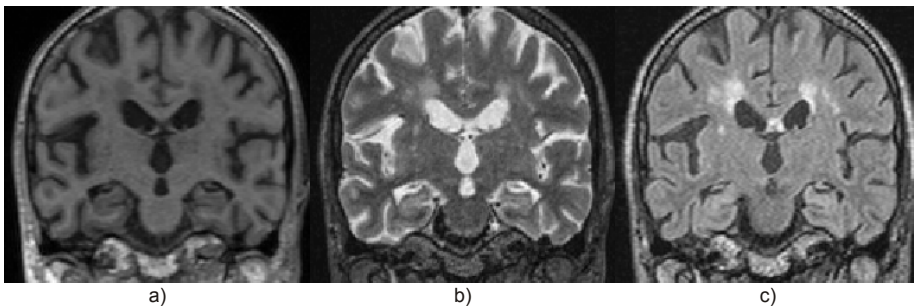
Figure 1.2: Coronal slices of a subject's MR images: a) T1, b) T2 and c) FLAIR.

Other MRI modalities include perfusion- and diffusion-weighted imaging. The first analyzes the blood flow patterns to the brain tissues and can therefore detect microvascular perfusion abnormalities. This is particularly relevant considering that vascular factors have also shown to be involved in AD [22]. Diffusion-weighted imaging is based on the microscopic motion of water molecules (diffusion) in structurally anisotropic tissues, such as the bundles of neuron axons in the white matter, and is therefore sensitive to the presence of microstructural white matter impairments [23]. Despite being more sensitive to changes at a lower scale and at a more functional level, these techniques are not yet widespread in the clinical practice. However, recent studies point to the advantage of combining the three types of MRI modalities (structural, perfusion-weighted and diffusion-weighted) to help understand the processes underlying the development of AD [24].

1.3   Structural MRI biomarkers to detect early-stage AD

Structural MRI (particularly T1-weighted imaging) has shown the presence of groupwise differences between healthy controls and (early) AD patients, mostly in medial temporal structures like the hippocampus and the entorhinal cortex. However, these studies have limited diagnostic value, since they only focus on global group differences. More recently, and with the development of machine learning techniques capable of dealing with high-dimensional data, methods have been proposed that perform classification between normal elderly controls and early-stage AD, having thus the potential to provide a diagnosis [25].

In this section, we briefly review such methods. We subdivide them into four categories, according to the type of features considered: 1) volumetric (features such as the volume and/or the shape of specific brain structures); 2) morphometric (voxelwise features, obtained after a non-linear registration to a template); 3) textural (image texture descriptors, determined both within specific brain structures or in the entire brain); 4) white matter hyperintensities descriptors (volume, spatial location and textures of perfusion-related lesions in the white matter).

1.3.1   Volumetric

Atrophy in medial temporal structures, such as the hippocampus and entorhinal cortex, has been considered a valid MRI biomarker of AD [9].

Methods have been proposed that use the hippocampal volume [26, 27] or its shape [28] as features in the classification of MCI. The volumes of the entorhinal cortex [29] and the amygdala [30] have also been considered for the same purpose. Similarly, the shape and the volume of the brain ventricles have recently shown promising results in the classification of MCI [31]. Finally, cortical thickness has also been considered in the early detection of AD [32].

However, it has been shown that medial temporal atrophy alone lacks specificity to confidently diagnose AD and it has been suggested that other brain regions should also be considered [9]. Additionally, the progression of AD pathology has a complex pattern. It starts in medial temporal structures like the hippocampus and the entorhinal cortex, and subsequently spreads through most of the temporal lobe and the posterior cingulate, ultimately reaching the cortical regions. Therefore, measuring volumes of specific brain

regions of interest (ROIs) is likely to miss important information that is available in the three-dimensional MR image. Also, the AD atrophy pattern does not necessarily follow pre-determined anatomical boundaries [33].

Furthermore, such volumetric measurements, besides making *a priori* assumptions about the expectedly affected brain structures, require the segmentation of these structures from the MR images, which is a complex and, in the case it is performed manually, time-consuming task. In particular, although several automatic hippocampus segmentation methods have been proposed, they show significant variability in the measurement of atrophy rates due to differences both in the methodological approaches and especially in the definition of the hippocampal boundaries [34]. On the other hand, manual segmentations of the hippocampus by experienced neuroradiologists suffer from intra- and inter-rater variability. They are also subject to the definition of anatomical landmarks, for which there is not yet a consensus [35].

1.3.2 Morphometric

Morphometric approaches comprise two main steps: non-linearly registering the brain images of all subjects to a common template and computing voxelwise measurements of interest. By statistically analyzing these voxelwise measures, it is possible to determine which voxels are significantly different between the subject groups, and maps showing the brain regions that are related to the disease can be created [36].

Furthermore, these voxel-by-voxel measurements can be taken as features, which are then fed to classifiers known to handle well high-dimensional data, such as Support Vector Machines (SVM), to discriminate between normal controls and early-stage AD [37, 25].

A widely used morphometric approach is to extract the voxelwise probability of the three brain tissues (cerebrospinal fluid, white and gray matter) that result from the fuzzy segmentation step performed prior to the non-linear registration to the template. This technique is called Voxel-Based Morphometry (VBM) [38]. In particular, the gray matter probability map (often referred to as "density" or "concentration" map) is the most often used, based on the assumption that AD primarily affects the cortical structures, as a consequence of the underlying neuronal loss [37]. Similarly, Deformation-Based Morphometry (DBM) considers the properties of the deformation field that results from the non-linear registration step [39, 40]. In particular, Tensor-Based Morphometry (TBM) is a variant of DBM and uses the voxelwise Jacobian determinant of this deformation field. This measure represents the change in volume that a voxel

undergoes during the non-linear registration and is therefore an indicator of local volume differences [41].

An advantage of these methods, with respect to the above-mentioned ROI-based volumetric approaches, is the fact that they do not require *a priori* assumptions about the size, location or number of regions to be analyzed, since they provide voxelwise measures determined in the entire brain.

However, and as mentioned above, these approaches always require non-linear alignments to a template, in order to achieve voxelwise inter-subject correspondence. A drawback is that, due to the high anatomical variability of brain structures like the cortical folds, the non-linear registration in those regions is not straightforward and severe misalignments may occur, compromising the subsequent analyses [42]. Also, the quality of the alignment is difficult to evaluate [43]. Finally, while non-linear registration can give more precise registration results than, for example, affine registration, there is also the risk of an over-alignment which can result in the elimination of informative patterns from the images [43].

1.3.3   Texture analysis

Volumetric and morphometric studies rely on large-scale structural alterations, such as volume/shape changes, and therefore only indirectly measure the changes that are known to occur, in AD, at the cellular level. Furthermore, these macroscopical alterations occur mostly at later stages of the disease, when neurodegeneration has already taken place [9].

The T1 intensities have been shown to be sensitive to degenerative age changes in the white matter [44]. Other studies show that T2 hypointensities are also present in AD brains [45]. The analysis of the MRI signal (intensities) may therefore bring additional information to the early diagnosis of AD that is otherwise missed by the structural-based volumetric/morphometric approaches. Furthermore, the local composition of brain tissues is also reflected in the MR intensity distribution, meaning that, for example, locally shrunk brain structures will display a different proportion of gray matter and cerebrospinal fluid compared with when they are unaffected.

Texture analysis is an image processing tool that has recently found applications in the study of various neurological diseases, including AD [46, 47, 48]. It extracts information that is not visible by a direct analysis of the image intensity and shape properties. In particular, a 3D MR image is a collection of volume elements (voxels), which are characterized by spatial locations and gray level intensities. Texture analysis evaluates the organizational pattern of these

voxels. The extracted features reflect the structure of the imaged tissues. Intuitive textural properties of an image include smoothness/roughness, regularity/irregularity, fineness/coarseness [49, 50].

In [48], the authors perform 2D texture analysis using the entire brain to classify between AD and normal controls. Rajeesh *et al* performed a similar study, but computed the textures only at the hippocampi [51]. In [52], Zhang *et al*. discriminate between normal controls and AD using 3D texture features computed at manually defined spherical ROIs, in the hippocampus and the entorhinal cortex. However, the results vary significantly with the location and the size of the chosen ROI, with accuracies ranging from 64% to 96%. Furthermore, in neither of these two studies is an analysis with MCI/early-stage AD subjects performed. Other studies have carried out texture analysis in the corpus callosum and thalamus, but the focus is on groupwise analyses rather than the classification of individual subjects [53].

The texture descriptors used in the above-mentioned studies have been computed at manually segmented ROIs, thus suffering from the same drawbacks as the volumetric approaches described above - they require *a priori* knowledge about the disease and depend on the quality of the segmentations. Also, although seemingly promising, texture analysis has not been thoroughly explored in the field of early detection of AD. In particular, to the best of our knowledge, no comparisons have been performed between the performance of texture and volumetric/morphometric descriptors in the classification of early-stage AD and the question about the usefulness of such image descriptors remains open.

1.3.4   White matter hyperintensities

White matter hyperintensities (WMH) are diffuse white matter abnormalities that are often associated with chronic cerebral ischemia, in particular with microvascular lesions originated by small vessel atherosclerosis [54]. They occur often in the elderly [55, 56, 57, 58] and have been shown to predict an increased risk of stroke, cognitive decline and death [59].

In structural MRI, WMH show as hypointensities in T1-weighted images and hyperintensities in T2-weighted images and Fluid Attenuated Inversion Recovery (FLAIR). In Figure 1.2, examples of such lesions can be observed (more clearly in the FLAIR image because of its contrast properties). Causes for signal changes in the lesions include demyelinization, axonal loss, gliosis, or edema [60].

The most widely used techniques to assess WMH in structural MR images are based on the lesion segmentation binary results. In particular, both the volume and the spatial distribution of these lesions have been thoroughly investigated. A recent long-term longitudinal study shows that the WMH load (volume) increases rapidly in normal subjects who develop MCI a decade later, suggesting that these lesions might serve as very early biomarkers of MCI and thus help with the earlier detection of AD [61]. In [62], the authors observe that the total lesion volume with a high proportion of lesions in the temporal region is associated with the risk of developing MCI.

However, there is still some controversy regarding the actual role played by WMH in the development of MCI and AD, with some studies showing that there is no relation between lesion volume and cognitive decline [63, 64, 65].

While the majority of such studies is based on volumetric analyses, more recently diffusion- and perfusion-weighted MR images have also been used to analyze microstructural properties of the WMH that go beyond their volume and location. In [66] the authors analyzed lesion perfusion differences between a group of normal controls and AD patients. They observed that the WMH locations were less perfused in AD than in the healthy subjects.

WMH also seem, then, to contain information that can help diagnose AD at an early stage. However, not much research has been done that uses such information to perform classification. In [67], the authors use texture descriptors, determined at the lesion locations in the images, to classify between a group of normal controls and a group of patients suffering from dementia of various types, including AD. Their results indicate a higher discriminative power of the lesion textures compared to structural properties like their volumes and locations.

1.4 Research scope and objectives

In this thesis, our goal is to detect Alzheimer's disease at an early stage of development, using structural MR images.

In order to achieve that, we use machine learning techniques to classify between groups of cognitively healthy controls and early-stage AD. In particular, we focus on feature extraction approaches that are based on texture analysis and that do not require prior knowledge nor segmentations of expectedly affected brain structures. The exception is Chapter 3, where we perform texture analysis on previously segmented white matter hyperintensities, to independently evaluate whether these lesions contain discriminative information.

Furthermore, we limit our scope to cross-sectional studies, in which we analyse groups of normal controls and early-stage AD subjects at one time instant, to investigate the possibility to provide a diagnosis based on a subject's single MR acquisition.

Specifically, the classification methods we propose are tested in groups of elderly controls *vs*. MCI (Chapters 3 and 5) and elderly controls *vs*. very mild to mild AD (Chapters 4 and 6). The terminology is based on the information available at the database from which the data is retrieved. As explained above, the boundaries between MCI and very mild AD are not objective nor consensual, and the two groups often overlap. However, this issue is out of the scope of this work. Also, the ground truth we use in the classifications is based on the clinical diagnosis (no pathological confirmation of AD was available).

We address the following research questions:

- Can we accurately segment white matter hyperintensities from a single MRI modality (FLAIR)? *(Chapter 2)*

- Is the proposed method comparable, in terms of performance, to existing multimodal approaches? *(Chapter 2)*

- Is it possible to detect MCI using only textural properties of white matter hyperintensities and what is the performance in comparison with volumetric/spatial features? *(Chapter 3)*

- Do intensity histograms contain enough information to detect early-stage AD and how do they perform in both traditional and dissimilarity-based classification frameworks? *(Chapter 4)*

- Can texture features help to classify between normal controls and MCI/early-stage AD and how to they perform compared to structural-based features? *(Chapters 5 and 6)*

- Can local patches help in both the classification of MCI/early-stage AD and the localization of the affected brain structures? *(Chapters 4, 5 and 6)*

- Do the detected regions correspond to what is already known about the affected brain structures? *(Chapters 4, 5 and 6)*

1.5 Thesis outline

The remainder of this thesis is organized as follows. In Chapter 2, we propose an automatic segmentation method for white matter hyperintensities that uses only three-dimensional FLAIR images and compare it with state-of-the-art approaches that use at least two modalities. Chapter 3 presents a classification method based on texture analysis of these lesions to discriminate between normal controls and MCI, using T1, T2 and FLAIR images. In Chapter 4, we propose a dissimilarity-based classification approach that uses simple image histograms both globally at the whole brain and locally at small patches to classify early-stage AD. In Chapter 5, we determine local texture maps of the whole brain and use them to detect MCI. Subsequently, in Chapter 6 we further use texture analysis on local image patches to both classify and localize AD at an early stage of development. Finally, Chapter 7 concludes this thesis by summarizing the research results and providing recommendations for future work.

Segmentation of white matter hyperintensities in FLAIR images

In the previous chapter, we have reviewed the state-of-the-art in the study of early-stage Alzheimer's disease using Magnetic Resonance images. One current line of research concerns the study of white matter hyperintensities and their possible role in the development of Alzheimer's disease. In this chapter, we propose an automatic segmentation method of white matter hyperintensities that requires only one Magnetic Resonance Imaging modality (FLAIR) and that can, therefore, be suitable for large-scale clinical trials. We evaluate it against the manual segmentation by a neuroradiologist and compare it, in a benchmark dataset, with more complex state-of-the-art multimodal methods.

This chapter is based on the following publication: *Lopes Simoes, A.R. and Moenninghoff, C. and Wanke, I. and Dlugaj, M. and Weimar, C. and van Cappellen van Walsum, A. and Slump, C.H., Automatic segmentation of cerebral white matter hyperintensities using only 3D FLAIR images. Magnetic Resonance Imaging, vol. 31, no. 7, pp. 1182-1189, 2013.*

2.1   Abstract

Magnetic Resonance (MR) white matter hyperintensities have been shown to predict an increased risk of developing cognitive decline. However, their actual role in the conversion to dementia is still not fully understood. Automatic segmentation methods can help in the screening and monitoring of Mild Cognitive Impairment patients who take part in large population-based studies. Most existing segmentation approaches use multimodal MR images. However, multiple acquisitions represent a limitation in terms of both patient comfort and computational complexity of the algorithms. In this work, we propose an

automatic lesion segmentation method that uses only three-dimensional Fluid-Attenuation Inversion Recovery (FLAIR) images. We use a modified context-sensitive Gaussian Mixture Model to determine voxel class probabilities, followed by correction of FLAIR artifacts. We evaluate the method against the manual segmentation performed by an experienced neuroradiologist and compare the results with other unimodal segmentation approaches. Finally, we apply our method to the segmentation of Multiple Sclerosis lesions by using a publicly available benchmark dataset. Results show a similar performance to other state-of-the-art multimodal methods, as well as to the human rater.

2.2   Introduction

White matter hyperintensities (WMHs) are diffuse white matter abnormalities that appear with high intensities in T2-weighted Magnetic Resonance (MR) images. Although the pathogenesis of WMHs is not yet completely understood, these lesions are often associated with chronic cerebral ischemia, in particular with microvascular lesions originated by small vessel atherosclerosis [54]. They occur often in the elderly [55, 56, 57, 58] and have been shown to predict an increased risk of stroke, cognitive decline and death [59].

The analysis of the real influence of WMHs on the development of dementia requires clinical studies involving large patient cohorts. Also, an accurate description of the location, shape and volume of the WMHs is necessary. Typically, WMHs are classified according to visual scales, such as the Scheltens scale or the Fazekas scale [68]. However, the results obtained by these visual scales are seldom comparable [69]. In addition, they have been shown to be little sensitive to clinical group differences [70]. Finally, they offer only a qualitative description of the WMHs, originating high intra- and inter-subject variabilities [71].

A quantitative and more reliable way of assessing WMHs is by manually determining the lesion volumes. However, for three-dimensional data this typically requires a slice-by-slice analysis, making the whole process cumbersome and time-consuming for the neuroradiologist. Also, the intra- and inter-rater variability have been reported to be high [72]. Clinical studies with hundreds of patients require, therefore, automated and robust segmentation methods.

Several methods have been proposed to automatically segment WMHs from MRI images, most of them using various types of MRI modalities [73, 74, 75]. The use of multimodal data presents several disadvantages. Namely, the acquired datasets must be coregistered, making the segmentations

computationally intensive and more prone to errors. In particular, motion artifacts are seen frequently in the MRI data from elderly patients, who are often not able to lie still during the whole acquisition period. This represents a serious limitation for the registration algorithms and can negatively influence the outcomes [76, 77].

Other methods have been specifically designed to segment Multiple Sclerosis (MS) lesions [78, 79]. Although MS lesions look similar to vascular-related WMHs in MR images, the spatial distribution of the lesions is often very different, with MS lesions occurring commonly in the corpus callosum and being symmetrically distributed in the brain, unlike the vascular WMHs [80].

WMHs are characterized by a larger T2 relaxation rate due to increased tissue water content and degradation of myelin [76]. Fluid-attenuated inversion-recovery (FLAIR) is a T2-weighted MR modality in which the cerebrospinal fluid (CSF) signal is attenuated. In FLAIR images, WMHs are characterized by an intensity range that only partially overlaps with that of normal brain regions, making this MRI modality well suited for lesion segmentation purposes [81].

Despite being the preferred imaging modality used by neuroradiologists to assess WMHs in the clinical setting, FLAIR has seldom been used alone in the automatic detection of these lesions [76, 77].

In [76], the authors determined an optimal FLAIR intensity threshold to separate WMHs from normal brain tissue, based on the analysis of the image histograms on a training set. More recently, Ong *et al*. [82] have applied an outlier detection approach to find this optimal threshold, followed by a false positive correction step that uses the co-registered T1-weighted image. Similarly, de Boer *et al*. [75] determined the optimal intensity threshold on a training set and used the T1-weighted image to ensure the detected lesions were all within the white matter.

Applying a threshold allows only for crisp segmentation and does not account for the Partial Volume Averaging (PVA) effect that is present in MR images. Having that in mind, Khademi *et al*. have proposed a segmentation method that allows for fuzzy segmentation and is based on a PVA model in FLAIR images [77].

In the methods described above, only the voxel intensity information is considered. However, it has been recognized that this makes methods highly sensitive to noise. In particular, boundary detection becomes problematic in noisy images. Furthermore, the common assumption that the voxel intensities are independent does not hold in practice. In reality, and intuitively, we can expect a certain voxel's value to be affected by those in its neighborhood [83, 84].

In this work, we propose a WMH segmentation method that uses solely FLAIR images. It is based on a modified Gaussian Mixture Model (GMM) that incorporates neighborhood information, followed by a false positive correction step, where common FLAIR artifacts [85] are eliminated from the segmentation.

Gaussian Mixture Models (GMM), estimated by the Expectation-Maximization (EM) algorithm, have been widely used in brain image segmentation [86, 87]. They provide a statistical description of the voxels' intensities and allow for fuzzy classification [88]. Because the traditional GMM-EM method is based only on intensity information, we use a modified GMM-EM method, initially proposed in [84], that considers additional contextual information. All initialization parameters are derived from the FLAIR image histogram.

We compare the performance of the proposed method with other unimodal approaches. For each method, the optimal parameters are determined using a training set that is retrieved randomly from our patient database. Evaluation is performed using the remaining patient datasets against the manual segmentation performed by an experienced neuroradiologist. Finally, we apply the method to a publicly available dataset of MS patients and compare the obtained performance results with those by multimodal segmentation methods and with the human expert.

2.3   Methods

Figure 2.1 shows the general overview of our method.

The raw FLAIR image is first preprocessed to remove the skull and to correct for bias field inhomogeneities. Subsequently, a context-sensitive GMM is applied to the brain image and the resulting WMH probability class is thresholded. Finally, the existing FLAIR artifacts (located at the interface between the cerebrospinal fluid and the gray matter and inside the ventricles - red pixels in the last figure) are eliminated by morphological processing of the cerebrospinal fluid segmentation mask, resulting in the final segmentation of the WMH (blue pixels in Figure 2.1d)). In the following subsections we will describe these steps in detail.

2.3.1   Gaussian Mixture Model

Figure 2.2 shows the histograms of the FLAIR images of two patients. Two peaks can be easily distinguished: the one at lower intensities corresponds to
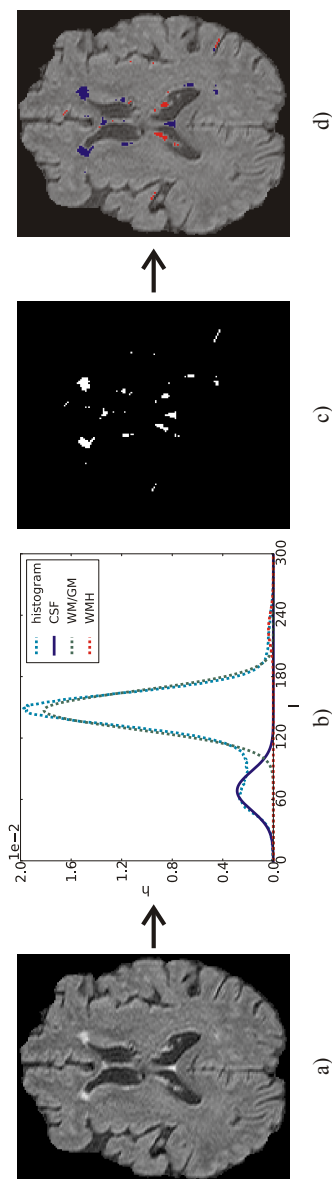
Figure 2.1: General overview of the segmentation method: a) we take the histogram of the skull-stripped and bias field-corrected FLAIR image and b) fit a 3-class context-sensitive GMM to it. Subsequently, we apply a threshold to the WMH class probability map, obtaining c) an initial lesion segmentation. Finally, we apply a post-processing step that corrects for artifacts in the initial segmentation; d) in red, the removed artifacts; in blue, the final segmentation. (For interpretation of the color references, we refer the reader to the web version of this article.)

cerebrospinal fluid voxels; the highest peak refers to white and gray matter voxels. Additionally, in Figure 2.2b) a low and broad peak is present at the right-end tail of the histogram. This peak is especially prominent in patients with a large lesion load and corresponds to WMH intensities.

We assume that the data can be modelled by a Gaussian Mixture Model (GMM) and that each voxel belongs to one of three distinct classes—cerebrospinal fluid (CSF), white and gray matter (WM/GM), or white matter hyperintensity (WMH)—. The probability density function (pdf) of a gray-level x can then be described by:

$$p(x|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{k=1}^{3} \pi_k \mathcal{N}(x|\mu_k, \sigma_k) \tag{2.1}$$

with $k = 1, 2, 3$ respectively corresponding to the CSF, WM/GM and WMH classes. Each Gaussian component $\mathcal{N}$ is characterized by a mixing weight π_k, a mean value μ_k and a standard deviation σ_k. We use the Expectation-Maximization (EM) algorithm to find these parameters.

Traditional Expectation-Maximization

The EM algorithm is an iterative procedure that maximizes the log-likelihood of the parameters [89, 90]. It alternates between two consecutive steps: the Expectation (E)-step and the Maximization (M)-step. In the E-step, the parameters at the current iteration are used to compute the log-likelihood. In the M-step, the computed log-likelihood is maximized to determine the new parameters.

Assuming that the data, $\mathbf{X} = (x_1, ..., x_N)$, are independent and identically distributed variables, the log-likelihood of the parameters given the data is defined as:

$$\ell(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}|\mathbf{X}) = \log \prod_{n=1}^{N} p(x_n|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{n=1}^{N} \log p(x_n|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}) \tag{2.2}$$

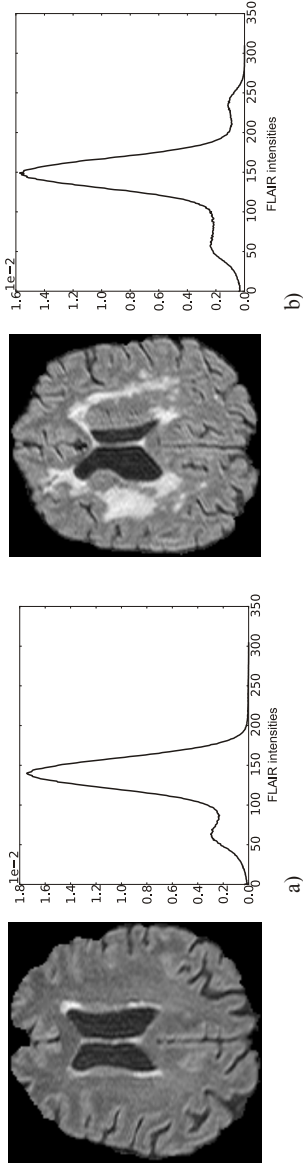The M-step parameter estimates are derived by maximizing Eq. (2.2):

Figure 2.2: FLAIR image and respective histogram from a patient: a) with a low WMH load; b) with a high WMH load.

$$\mu_k^{(i+1)} = \frac{1}{N} \sum_{n=1}^{N} x_n T_{k,n}^{(i)}$$

$$\sigma_k^{(i+1)} = \sqrt{\frac{\sum_{n=1}^{N} \left(x_n - \mu_k^{(i+1)} \right)^2 T_{k,n}^{(i)}}{\sum_{n=1}^{N} T_{k,n}^{(i)}}} \qquad (2.3)$$

$$\pi_k^{(i+1)} = \frac{1}{N} \sum_{n=1}^{N} T_{k,n}^{(i)}$$

where $T_{k,n}^{(i)}$ is determined at the E-step by:

$$T_{k,n}^{(i)} = \frac{\pi_k^{(i)} \mathcal{N}(x_n | \mu_k^{(i)}, \sigma_k^{(i)})}{p(x_n | \boldsymbol{\pi}^{(i)}, \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{(i)})} \qquad (2.4)$$

The initial parameters are computed from the histogram as follows: $\mu_{WM/GM}^{(0)}$ and $\mu_{CSF}^{(0)}$ correspond to the first and second highest peaks in the histogram, respectively; $\mu_{WMH}^{(0)}$ is taken as the local histogram maximum between $\mu_{WM/GM}^{(0)}$ and the maximum intensity (if no local maxima are found, we take this value as the average between $\mu_{WM/GM}^{(0)}$ and the maximum intensity); all standard deviations are initialized with the same value: the standard deviation of the voxel intensities in the CSF class (with the threshold for this class being the local minimum between $\mu_{WM/GM}^{(0)}$ and $\mu_{CSF}^{(0)}$); finally, the initial class weights are selected based on the relative ratios between $\mu_{WM/GM}^{(0)}$, $\mu_{CSF}^{(0)}$ and $\mu_{WMH}^{(0)}$. These weights can take values in the interval [0,1]. This means that if there are no lesions in the brain the outcome will be a two-class segmentation (CSF and WM/GM).

The algorithm has converged when the absolute normalized difference between the log-likelihood values at two consecutive iterations is lower than tolerance $T = 10^{-3}$.

Although it may be sufficient to obtain a first rough approximation of the voxels' statistical distributions, the traditional GMM-EM algorithm has the disadvantage of taking only intensity information into account. We therefore apply a previously proposed [84] adaptation to the E-step. The difference between the performance of the normal and the modified GMM-EM approaches is particularly significant in images with low WMH loads, as we will show in Section 2.4.

Context-Sensitive Expectation-Maximization

In [84], the authors introduced contextual information into the traditional GMM-EM method as follows. At each iteration, the posterior probability (Eq. (2.4)) is substituted by:

$$T_{k,n}^{(i)CC} = \frac{\pi_k^{(i)} C_{k,n}^{(i)} \mathcal{N}(x_n|\mu_k^{(i)}, \sigma_k^{(i)})}{p(x_n|\boldsymbol{\pi}^{(i)}, \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{(i)})}, \tag{2.5}$$

which incorporates a context-sensitive penalty term $C_{k,n}^{(i)}$. This term imposes that, at each iteration, the probability that a voxel belongs to class k depends not only on the voxel's intensity, but also on its neighbors' current class probabilities. We define the penalty term as follows:

$$C_{k,n}^{(i)} = \Phi\{I_k^{(i)}\}(x_n) \tag{2.6}$$

with $I_k^{(i)}$ being the membership image which, at each brain voxel x_n, represents the probability that the voxel belongs to class k. $\Phi\{\cdot\}$ represents the filter used to take the voxel's neighborhood into account.

We initialize the context-sensitive (CS-) EM method with the parameters that result from applying the traditional GMM-EM method to the dataset. After convergence, we apply thresholds t_{WMH} and t_{CSF} to the resulting WMH and CSF membership images, respectively.

2.3.2 False Positive correction

After applying the threshold to the WMH probability map, we still obtain some false positives — voxels that are initially considered to be lesions but are in reality FLAIR artifacts. We apply a postprocessing step that consists of eliminating these voxels from the segmentation.

A common location of false positives is in the interface between the CSF and the cortical gray matter. To eliminate these voxels from our initial segmentation, we use the CSF mask obtained after thresholding the CSF class membership image that results from the segmentation method described above. We perform binary dilation of this mask with a three-dimensional cubic structure with size S×S×S. We mask our first WMH segmentation obtained after applying the EM method with the dilated CSF mask.

Other hyperintense voxels, resulting from flow artifacts (located mainly in the ventricular system) [85] are also eliminated in this step by morphologically "closing the holes" [91] in the dilated CSF mask.

Finally, and because the lesion voxels adjacent to the ventricles are also eliminated after this step, we perform binary propagation [91] to the initial WMH segmentation in order to recover these wrongly eliminated voxels.

2.3.3   Evaluation metrics

To evaluate the method, we compare our results with the manual segmentation provided by an experienced neuroradiologist. We use the following metrics for comparison: Dice Similarity Coefficient (DSC), Overlap Fraction (OF) and Extra Fraction (EF) [73]:

$$\text{DSC} = \frac{2 \times \#\text{TP}}{\#\text{AS} + \#\text{GT}} \tag{2.7}$$

$$\text{OF} = \frac{\#\text{TP}}{\#\text{GT}} \tag{2.8}$$

$$\text{EF} = \frac{\#\text{FP}}{\#\text{GT}} \tag{2.9}$$

with TP and FP being the true and the false positives, respectively, AS the automatic segmentation and GT the ground truth provided by the expert.

Because the lesion load (LL) is often an important measure in clinical studies, we finally determine the correlation coefficient between the obtained LL values with those from the manual segmentations.

2.4   Experiments and Results

2.4.1   Data

Forty datasets were retrieved from a large database of a cognition study with MCI and control subjects carried out at the University Hospital of Essen, Germany. From these 40 subjects, 15 correspond to stable normal controls, 14 to stable amnestic-MCI subjects, 8 to MCI subjects who have progressed to dementia and 3 to normal subjects who have declined to amnestic-MCI. The age of the subjects is 74.7 ± 4.3 (range 62-82).

Three-dimensional isotropic FLAIR images are utilized in this study (1.5 T Siemens Avanto, Germany); TR = 6000ms; TE = 308ms; TI = 2200ms; voxel size = 1mm^3). We apply the following preprocessing steps to the raw FLAIR images:

- brain extraction using BET (FMRIB's Brain Extraction Tool, `http://fsl.fmrib.ox.ac.uk/fsl/bet2`) [92];
- bias field correction using FAST (FMRIB's Automated Segmentation Tool, `http://fsl.fmrib.ox.ac.uk/fsl/fast4`) [93]

For the evaluation of the method, we use as the ground truth the manual segmentation performed on all 40 FLAIR images by an experienced neuroradiologist using 3D Slicer (www.slicer.org).

The WMH lesion loads are typically divided into three groups: low LL (less than 10 cm^3), medium LL (between 10 and 30 cm^3) and large LL (more than 30 cm^3). After manual labeling, we obtain 18 datasets that are considered to have low LL, 13 datasets with medium LL and only 9 datasets with high LL.

We randomly split our dataset into 30% training and 70% test. That is, we use 12 datasets (four of each LL category) to learn our method's optimal parameters, while the remaining 28 datasets are used as a test set for an independent evaluation of the method.

2.4.2 Selection of the optimal parameters

First WMH segmentation

Two parameters influence the outcome of the first step of the segmentation method: the threshold which is applied to the WMH class membership to obtain a crisp segmentation and the neighborhood filter type and size ($\Phi\{.\}$ in Eq. (2.6)).

We use the training set to find the optimal joint parameters. Figure 2.3 shows the joint parameter analysis - on the horizontal axes, we plot the threshold values and the filter types. The z-direction shows the corresponding DSC values averaged across the training set. We observe that the DSC index is most sensitive to t_{WMH}, with very little variability across the various neighborhood types. At the optimal threshold (10^{-5}), the average DSC values vary less than 5% across the considered neighborhood types. The exception is the case where no neighborhood information is used. This approach, as we will also show in Section 2.4, performs considerably worse than the contextual methods.

We then select the first neighborhood (the $3 \times 3 \times 3$ mean filter) for further processing.

For this neighborhood filter, we plot each subject's DSC curve and the average across all training set subjects. The broader curve, with a lower optimal threshold, corresponds to a low LL dataset. On the other hand, the datasets with higher LL have higher optimal thresholds.
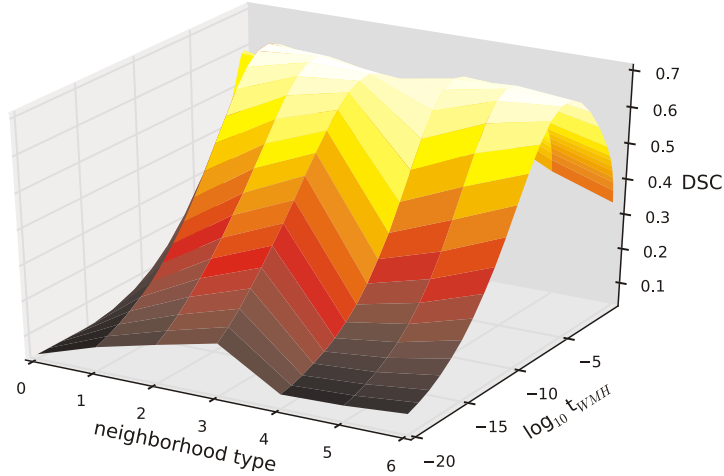
Figure 2.3: Search for the optimal parameters of the first step of the segmentation method. The neighborhood filter types are the following: 0: no neighborhood information (traditional GMM-EMM method); 1: mean filter with size 3×3×3; 2: mean filter with size 5×5×5; 3: mean filter with size 7×7×7; 4: isotropic Gaussian filter with $\sigma = 0.7$; 5: isotropic Gaussian filter with $\sigma = 1.5$; 6: isotropic Gaussian filter with $\sigma = 2$.

False positive correction

Finally, we correct for the presence of FLAIR artifacts. This step takes also two parameters: the threshold of the CSF membership image and the size of the structuring element used to create the FP mask from the CSF segmentation.

Similarly to what was done in the previous subsection, we analyze the joint parameters and select the combination that gives the best results on the training set. In this case, we fix the WMH threshold to 10^{-5} and the neighborhood filter to the mean in a 3×3×3 local window.

As in the previous case, the CSF threshold has the most influence on the DSC value, with the best performance being achieved at $t_{CSF} = 10^{-2}$ and with a structuring element size of $5 \times 5 \times 5$. However, for thresholds greater than 10^{-5}, the mean DSC values also vary less than 5%, regardless of the structuring element size.
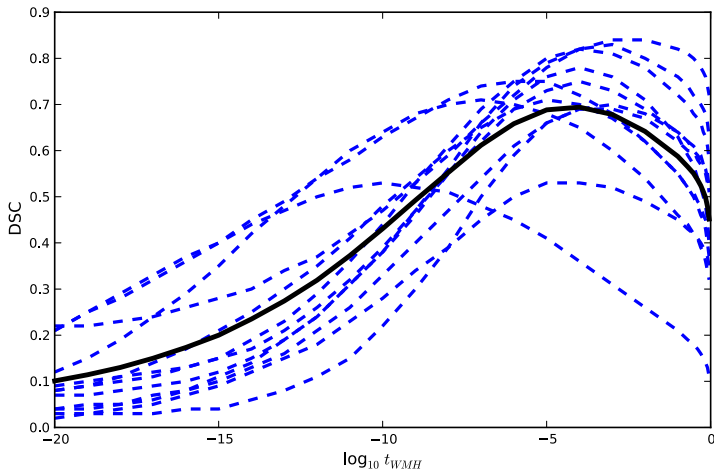
Figure 2.4: DSC values for all patients in the training set, using a mean filter with size 3×3×3. The average DSC corresponds to the thicker black line.
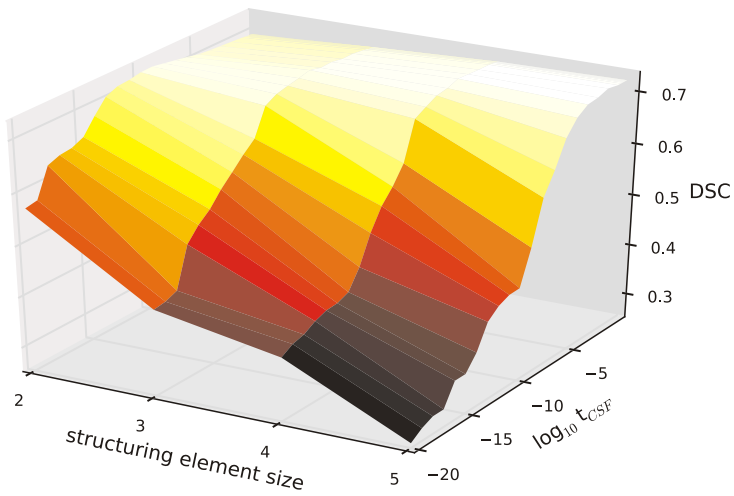

Figure 2.5: Search for the optimal parameters in the FP correction step.

2.4.3   Evaluation on the test set

We evaluate the method against the manual segmentation on the remaining 28 datasets. Table 2.1 shows the final DSC, EF and OF values, per lesion load, in the test set.

Table 2.1: Performance measures for the 28 patients in the test set.

LL category	Subject ID	DSC	EF	OF	AS (cm^3)	GT (cm^3)
Low	1	0.79	0.19	0.78	5.787	5.973
	2	0.44	0.10	0.31	2.048	4.887
	3	0.70	0.21	0.65	3.781	4.380
	4	0.60	0.11	0.47	3.353	5.742
	5	0.64	0.14	0.54	4.547	6.708
	6	0.21	0.01	0.12	1.177	9.168
	7	0.25	0.03	0.15	0.375	2.160
	8	0.37	0.01	0.23	1.170	4.896
	9	0.70	0.30	0.69	8.100	8.160
	10	0.49	0.03	0.34	1.322	3.583
	11	0.37	0.01	0.23	0.919	3.801
	12	0.40	0.05	0.26	0.698	2.226
	13	0.67	0.15	0.53	4.945	7.062
	14	0.51	0.27	0.43	0.497	0.714
Medium	15	0.72	0.28	0.72	10.291	10.267
	16	0.63	0.15	0.53	8.113	11.917
	17	0.71	0.11	0.61	7.541	10.328
	18	0.74	0.17	0.69	11.471	13.475
	19	0.70	0.28	0.69	11.108	11.497
	20	0.39	0.09	0.26	3.963	11.375
	21	0.77	0.18	0.74	10.877	11.801
	22	0.83	0.17	0.84	13.403	13.313
	23	0.80	0.20	0.79	12.999	13.109
High	24	0.85	0.29	0.96	155.220	124.177
	25	0.86	0.18	0.89	40.293	37.559
	26	0.84	0.23	0.89	56.411	50.679
	27	0.81	0.33	0.90	73.326	59.881
	28	0.83	0.20	0.84	47.226	45.177

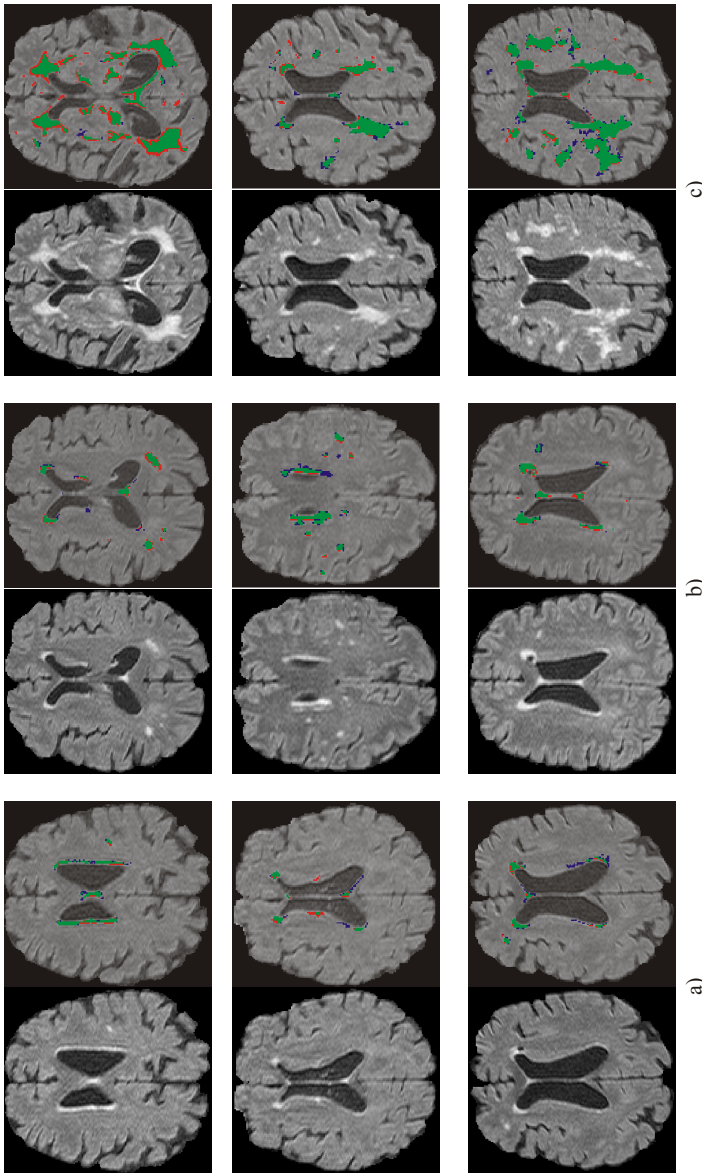The average DSC values are 0.51, 0.70 and 0.84 for the low LL, medium LL

Figure 2.6: Segmentation examples for the three lesion load categories: a) low, b) medium and c) high. Green: True positives; Red: False positives; Blue: False negatives.

and high LL, respectively. DSC values above 0.70 are considered to represent a very good agreement between segmentations [94]. The lower similarity values for the low lesion loads are to be expected, since errors in the segmentation have a greater impact on the similarity score when the lesion load is lower. This has also been reported in previous studies [74, 73, 95].

In Table 2.1 we can observe a systematic underestimation of the lesion loads in the low LL cases and an overestimation for the high LL datasets. The latter can be visualized on the first example of Figure 2.6c) and is also expressed on the relatively high EF values for the high LL datasets (Table 4.2).

Finally, we plot the automatically obtained LL against the ground truth LL (Figure 2.7). The obtained correlation coefficient ($R = 0.9966$) indicates a strong correlation between the two measurements.
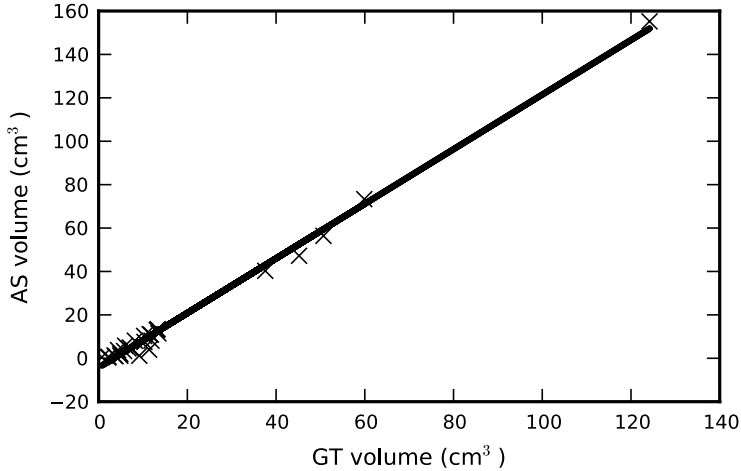
Figure 2.7: Ground Truth (GT) and Automatic Segmentation (AS) lesion loads and the fitted linear regression line ($y = 1.28x - 4.19$).

2.4.4   Comparison with other unimodal approaches

To further evaluate the performance of the proposed method, we compare it with four other segmentation approaches which use only FLAIR images. For each of these approaches, we search for the optimal parameters in the training set and evaluate them in the test set. The exception is the first method, in which

a threshold is applied to the FLAIR intensities (intensity thresholding, IT). In this case, because the goal is not to evaluate any specific method that searches for an optimal threshold, we take the optimal threshold value for each subject individually. This way we ensure that the obtained DSC is the highest that can be achieved with such approach.

The second comparison is with the traditional GMM, with parameters determined by EM (simple GMM, sGMM). This method, unlike the first one, yields a fuzzy segmentation. However, it is also based only on intensity information.

The PVA model introduced in [77] is used for the third comparison. Similarly to the GMM-EM method, its output is a fuzzy segmentation that does not consider any contextual information. However, this method is based not only on the image intensities but also on the gradient magnitudes.

Finally, we compare our approach with an analogous segmentation method - Fuzzy C-Means (FCM), modified in [96] to incorporate neighborhood information (cFCM). Unlike the GMM-EM approach we use here, this method does not assume any probabilistic model for the voxel intensities.

For the proposed method, we show the results obtained after the initial WMH segmentation ("proposed (first)") and after FP correction ("proposed (final)").

The results are shown in Table 2.2. Figure 2.8 shows the average DSC values obtained for the three LL categories.

We observe that the proposed method performs significantly better than the first three context-free approaches. A slight improvement is also observed with respect to the contextual FCM method. However, the FCM method seems to perform considerably less robustly in very low LL cases - particularly with respect to the EF measure.

In all cases, the DSC values are lower for the low LL cases. This is expectable, since errors in these measurements tend to have a larger impact on the final similarity score. Also, the variability is larger in these cases, indicating a lower robustness of the methods.

A criticism that can be made to model-based segmentation methods, such as GMM, is that, for low LL, there may not be enough lesion voxels to accurately derive the model's parameters [97]. Although this may be true for the simple GMM (with an average DSC of 0.38 in the low LL case), the problem seems to be overcome by considering contextual information, as in the proposed method, which outperforms the model-free contextual approach (cFCM).

It is worth noting that the performance of the first approach is highly over-

Table 2.2: Performance average (standard deviation) values for four different approaches and for the proposed method (first step and after FP correction).

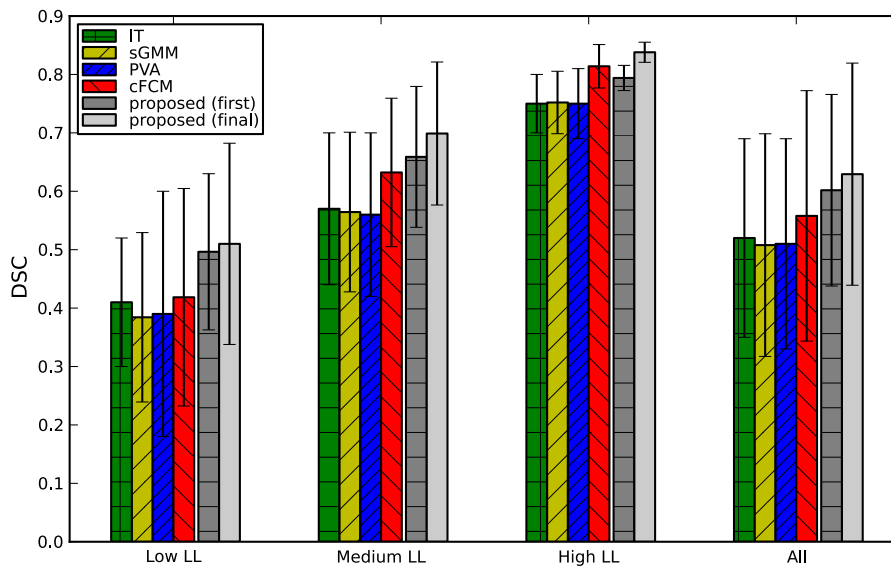| Methods	Low LL			Medium LL			High LL		
	DSC	EF	OF	DSC	EF	OF	DSC	EF	OF
IT	0.41 (0.11)	0.40 (0.11)	0.36 (0.10)	0.57 (0.13)	0.31 (0.17)	0.51 (0.12)	0.75 (0.05)	0.23 (0.07)	0.73 (0.07)
sGMM	0.38 (0.15)	1.0 (2.73)	0.34 (0.13)	0.56 (0.14)	0.14 (0.04)	0.46 (0.14)	0.75 (0.05)	0.15 (0.04)	0.70 (0.08)
PVA	0.40 (0.13)	0.40 (0.69)	0.32 (0.11)	0.56 (0.15)	0.11 (0.05)	0.45 (0.15)	0.75 (0.05)	0.19 (0.07)	0.71 (0.07)
cFCM	0.42 (0.19)	1.62 (2.99)	0.47 (0.12)	0.63 (0.13)	0.11 (0.06)	0.52 (0.13)	0.81 (0.04)	0.06 (0.02)	0.73 (0.06)
prop. (first)	0.50 (0.13)	0.36 (0.42)	0.46 (0.19)	0.66 (0.12)	0.34 (0.13)	0.67 (0.16)	0.79 (0.02)	0.37 (0.05)	0.90 (0.04)
prop. (final)	0.51 (0.17)	0.11 (0.09)	0.41 (0.20)	0.70 (0.12)	0.18 (0.07)	0.65 (0.16)	0.84 (0.02)	0.25 (0.06)	0.90 (0.04)

Figure 2.8: Average DSC values for the six compared methods, separated by lesion load.

estimated, since for each patient we take the optimal DSC value (without recurring to a training set). However, results also show that the two other context-free approaches (simple GMM and PVA model) have a similar performance, indicating that adding neighborhood information not only improves the similarity scores but also seems to be a determinant factor in the methods' performance.

Finally, a paired sample t-test on the results of all subjects on the test set shows a significant improvement ($p < 0.05$) on the DSC metric with the first step of the proposed method with respect to all other approaches. Furthermore, the second step also accounts for a significant improvement of the performance metrics with respect to the first step, indicating the importance of the artifact elimination step in the segmentation.

Table 2.3 shows the correlation coefficients between each segmentation approach and the manual measurements.

Table 2.3: Correlation coefficients between the lesion loads determined by the
automatic and the manual measurements.

IT	sGMM	PVA	cFCM	prop. (first)	prop. (final)
0.9969	0.9901	0.9927	0.9862	0.9957	0.9966

2.4.5   Robustness to the initialization parameters

A final evaluation is performed by varying the parameters that initialize the
first EM procedure. Converging to local minima is a well-known limitation of
the EM method [98]. Therefore, we evaluate the robustness of the proposed
method to variations in the three parameters of the Gaussian that describes the
WMH class distribution: the mean value μ_{WMH}, the standard deviation σ_{WMH}
and the weight π_{WMH}, determined as described in Section 2.3. Again, we use
the Dice Similarity Coefficient as a performance measure.

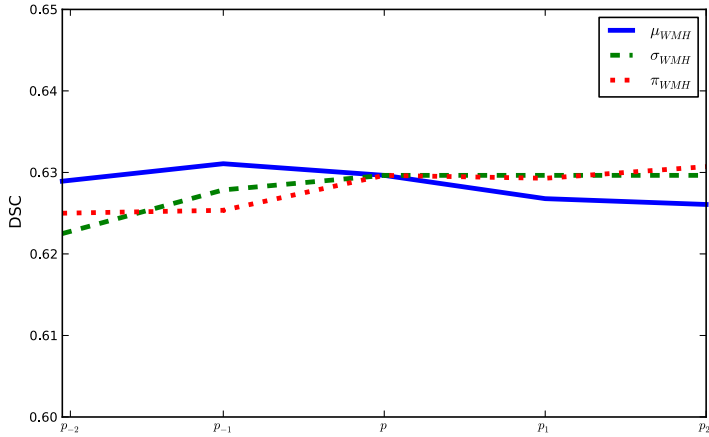The results are shown in Figure 2.9.

Figure 2.9: Variation of the average DSC values with varying initialization parameters.

In the horizontal axis we show the parameter values used for compar-
ison. During the evaluation of each parameter, the others remained con-
stant and equal to the values automatically determined by the method, as de-

scribed in Section 2.3. The values $\{p_{-2}, p_{-1}, p, p_1, p_2\}$ correspond to $\{\mu_{\text{WMH}} - 20, \mu_{\text{WMH}} - 10, \mu_{\text{WMH}}, \mu_{\text{WMH}} + 10, \mu_{\text{WMH}} + 20\}$ for the WMH mean, to $\{\sigma_{\text{WMH}} - 10, \sigma_{\text{WMH}} - 5, \sigma_{\text{WMH}}, \sigma_{\text{WMH}} + 5, \sigma_{\text{WMH}} + 10\}$ for the standard deviation and to $\{\pi_{\text{WMH}}/10, \pi_{\text{WMH}}/5, \pi_{\text{WMH}}, \pi_{\text{WMH}} \times 5, \pi_{\text{WMH}} \times 10\}$ for the WMH weight.

Even though we select a large range of parameter values, the DSC values remain approximately constant. For the mean value, the variability of the DSC scores (ratio between the range and the maximum value) is 0.7%. For the standard deviation and the class weight the variabilities are 1.1% and 0.9%, respectively.

2.4.6 Application in the segmentation of Multiple Sclerosis (MS) lesions

To show the applicability of our method in a different neurological disease, we use a benchmark dataset made available by the Medical Image Computing and Computer Aided Intervention Society's (MICCAI's) MS Lesion Segmentation Challenge 2008 (`http://www.ia.unc.edu/MSseg`). The data consist of 23 FLAIR images acquired at the Children's Hospital Boston (CHB) and at the University of North Carolina (UNC), with a dimension of $512 \times 512 \times 512$ voxels, resliced at 0.5 mm $\times$ 0.5 mm $\times$ 0.5 mm resolution using cubic spline interpolation.

The four error metrics used to evaluate the methods' performance are the following: relative absolute volume difference, average symmetric surface distance, true positive rate and false positive rate. The results were scaled to a range such that a score of 90 points is comparable to the performance of a human expert. For further details on the design of the Challenge, we refer the reader to [99].

The results for all subjects are shown in Table 2.4.

Our method obtained an overall score of 82.0055 (`http://www.ia.unc.edu/MSseg/results_table.php`), outperforming other WML segmentation methods in the literature [73, 82, 78] and reaching similar performance to other methods [79]. It is worth noting that our method performs less than 2 score points worse than the method that is currently at the first position of the Challenge. Also, all other participating methods require at least two MR modalities, while ours uses only FLAIR image data. Finally, some of the methods assume *a priori* knowledge about the spatial distribution of the MS lesions [100, 79]. In contrast, our method has a more general applicability since it uses only intensity information.

Table 2.4: Summary of the performance measures for the 23 patients in the MICCAI Challenge test set.

Ground Truth	UNC Rater								CHB Rater								
	Volume Diff.		Avg. Dist.		True Pos.		False Pos.		Volume Diff.		Avg. Dist.		True Pos.		False Pos.		Total
All Datasets	[%]	Score	[mm]	Score	[%]	Score	[%]	Score	[%]	Score	[mm]	Score	[%]	Score	[%]	Score	
Range	4.5-100	85-99	1.2-128	0-97	0-68.4	51-90	0-52	78-100	11.7-142.5	79-98	1.2-128	0-97	0-81.5	51-98	0-60.8	73-100	59-93
Std dev	37.7	5.6	42.8	36.5	21.3	12.2	20.6	9.2	35.3	5.3	42.6	35.4	23.4	13.4	20.0	9.3	11.9
Average	47.4	92.9	24.6	71.3	30.4	68.7	21.3	92.8	62.9	90.8	23.5	73.2	35.2	71.5	18.2	94.6	**82.0**

2.5 Conclusion

In this work, we present a method to automatically segment WMHs using only 3D FLAIR images. It uses a context-sensitive Gaussian Mixture Model to obtain class probabilities, followed by crisp segmentation and artifact correction. Unlike the majority of the existing approaches (to the best of our knowledge), our method requires no additional MRI modalities nor atlases, thereby shortening the acquisition time, avoiding the need for co-registrations and allowing for near real-time analysis. Results show that the method is suitable for a robust segmentation of WMHs of various loads. Also, a comparison with other segmentation approaches indicates the usefulness of, on the one hand, incorporating contextual information and, on the other hand, considering a model for the lesions (instead of a model-free approach such as FCM). The significant improvements observed on the performance measures after applying the FP correction step (with respect to the initial segmentation) suggest the efficacy of the simple CSF-based mask we have used, without needing additional MR modalities.

We have also demonstrated the applicability of our method in the detection of other lesion types, namely Multiple Sclerosis lesions. In particular, the results on a benchmark dataset show that our method performs comparably to other state-of-the art multimodal methods, with the difference that ours does not need any MR modalities other than FLAIR and does not make assumptions about the spatial distribution of the lesions, therefore having a wider applicability. The final score obtained in this evaluation indicates that the method performs close to the human observer. Because we make no assumptions about the lesion spatial distribution, we believe that this method can be applied to other neurological diseases that have a similar appearance in FLAIR images. Examples include subcortical arteriosclerotic encephalopathy and brain tumors.

A possible drawback of our method is that it requires two preprocessing steps: brain extraction and bias field correction. This is a consequence of the algorithm being fully intensity-based and relying on the brain image histogram. An extension can be considered in which the bias field correction is incorporated into the segmentation framework. Also, a study on the robustness of the method to the presence of field inhomogeneities and wrong brain extractions should be carried out. Finally, it is worth pointing out that we do not perform any registration step, which is typically more time-consuming than the two steps required by our method (particularly when using multimodal data).

Ultimately, we expect that this method can become a useful tool in the eval-

uation of WMHs in the large patient cohorts required by population-based studies.

CHAPTER 3

Texture analysis of white matter hyperintensities

In this chapter, we further analyse a number of properties of the white matter hyperintensities, which have been introduced in Chapter 2. In particular, we investigate the image textures at the lesions as possible early-stage Alzheimer's disease biomarkers. For that, we perform classification between a group of normal elderly controls and a group of Mild Cognitive Impairment subjects using local texture descriptors. We also assess whether texture descriptors are more discriminative in this classification task than the widely used lesion volumes and locations.

3.1   Introduction

Mild Cognitive Impairment (MCI) is considered to be a transitional stage between normal ageing and dementia, with impaired memory as first sign [12]. Some studies have demonstrated that MCI patients are at an increased risk of developing dementia when compared with healthy subjects of the same age group [101], while others claim that MCI generally represents very mild Alzheimer's Disease (AD) [102], with subjects progressing to more severe stages of dementia at different rates. In any case, MCI subjects are a group of interest in the development of techniques to detect early-stage AD.

Structural Magnetic Resonance Imaging (MRI) is a diagnostic tool that provides high-resolution images and a high brain tissue contrast. In addition, its non-invasiveness makes it a suitable imaging technique for follow-up studies. Most MRI studies on prodromal stages of AD focus on the large-scale analysis of gray matter structures. In particular, volume and shape changes at the level of the cortex [29], the hippocampi [27, 28] and the amygdalae [30], or the

cerebrospinal fluid (CSF) cavities such as the ventricles [103] have been widely explored as imaging markers of AD. However, recent studies have shown that the white matter is also affected at the early stages of the disease [104, 105].

White matter hyperintensities (WMH) are diffuse white matter abnormalities that are often associated with chronic cerebral ischemia, in particular with microvascular lesions originated by small vessel atherosclerosis [54]. They occur often in the elderly [55, 56, 57, 58] and have been shown to predict an increased risk of stroke, cognitive decline and death [59].

In structural magnetic resonance (MR) imaging, WMH show as hypointensities in T1-weighted images and hyperintensities in T2-weighted images and Fluid Attenuated Inversion Recovery (FLAIR). Causes for signal changes in the lesions include demyelinization, axonal loss, gliosis, or edema [60].

The most widely used techniques to assess WMH in structural MR images are based on the binary results of the lesion segmentation. In particular, both the volume and the spatial distribution of these lesions have been thoroughly investigated. A recent long-term longitudinal study shows that the WMH volume (clinically known as "lesion load") increases rapidly in normal subjects who develop MCI a decade later, suggesting that these lesions might serve as very early biomarkers of MCI and thus help towards earlier detection of Alzheimer's Disease (AD) [61]. In [62], the authors observe that the total lesion load with a high proportion of lesions in the temporal region is associated with the risk of developing MCI.

However, there is still some controversy regarding the actual role played by WMH in the development of MCI and AD, with some studies showing that there is no relation between lesion volume and cognitive decline [63, 64, 65].

While the majority of such studies is based on volumetric analyses, more recently diffusion- and perfusion-weighted MR images have also been used to analyze microstructural properties of the WMH that go beyond their volume and location. In [66], the authors analyze lesion perfusion differences between a group of normal controls (NC) and Alzheimer's disease (AD) patients. They observe that the WMH locations are less perfused in AD than in NC subjects.

Despite their higher sensitivity to microstructural white matter changes, diffusion- and perfusion-based imaging are less widespread in clinical trials than structural MR imaging. In particular, T1- and T2-weighted images are most often acquired. FLAIR, because of its contrast properties, is typically used to help detect and quantify WMH. Knowing that the presence of white matter lesions is reflected in signal changes in structural MR images, a pertinent question is whether these images contain sufficient information that can help characterize WMH and distinguish them between normal elderly and MCI.

Oppedal *et al*. [67] have extracted, from FLAIR images, 2D Local Binary Patterns (LBP), a texture descriptor that is widely used in computer vision, to classify between normal controls and patients with dementia. They determine statistical measures on the LBP label images (at the lesion voxels) and obtain higher classification rates than when using volumetric features.

To the best of our knowledge, Oppedal's study [67] is the only one that uses structural MR images to investigate textural differences, at the lesions' locations, between NC and dementia patients. Furthermore, they analyze these features' discriminative power by performing classification between the two groups.

However, the LBP descriptors they use are two-dimensional. Also, they analyze only a few statistical properties of the LBP labels (such as the mean, standard deviation, kurtosis, etc.) and not the entire histograms. Additionally, their feature analysis and combination is performed a posteriori, with the combination not including more than three selected features. Their analysis is also based on FLAIR images only. Finally, the dementia subject group includes not only Alzheimer's disease patients but also subjects with Lewy Body dementia [106].

In this work, we further explore the intensity content of three structural MR modalities (T1, T2 and FLAIR) at the lesions' locations. We investigate the use of 3D LBP descriptors to discriminate between a group of normal controls and MCI subjects. We incorporate both multiscale and multimodal information and investigate which features are most significantly different between the two groups. In particular, to avoid including large-scale structural differences such as enlarged ventricles and shrunk hippocampi, we use only small-scale texture descriptors. Finally, we compare the results with those obtained using volumetric/spatial location features and intensity histograms.

3.2   Methods

3.2.1   Local Binary Patterns

The local binary pattern (LBP) is an operator that labels the pixels of an image according to its local appearance, or small-scale structure. The most common texture descriptor based on the LBP operator is the histogram of the labels within a region of interest [107].

In its original form, proposed by [108], the LBP label of a certain pixel is determined by thresholding its 8 neighbors (in a 3×3 neighborhood) with its own

gray value and converting the resulting binary into a decimal number. More recently, several variants of the LBP descriptor have been proposed [107]. In particular, [109] extended the LBP descriptor to allow for other neighborhood configurations.

The formal description of such an operator is as follows. At a pixel with gray value g_c, with P neighbors at radius r, its LBP label is given by:

$$LBP_{P,r} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \qquad (3.1)$$

with

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \qquad (3.2)$$

Typical values of r are in the order of a few pixels. Accordingly, P takes normally values in the order of the number of neighbors for a certain r. Figure 3.1a) shows an example of a circular neighborhood. The points that, after bilinear interpolation of the respective neighborhood, have a lower intensity than the center pixel are shown in black and the others in white.

More recently, it was observed that the majority of the observed patterns are uniform (described by no more than two bitwise 0-1 transitions). For the 8-bit ($P = 8$) LBP descriptor, 58 out of the 256 possibilities correspond to uniform patterns. Furthermore, it is possible to obtain a rotation-invariant descriptor by rotating the LBP binary code into its minimum decimal value [110]. In general, for P neighbors we have $P + 1$ possible uniform rotation-invariant LBP labels (shown in Figure 3.1b)). All non-uniform patterns are assigned label $N = P + 2$.

Two extensions of the LBP operator to 3D data have been proposed in [111], mostly for spatio-temporal applications. In particular, LBP-TOP (Three Orthogonal Projections) is computationally simple. It consists of calculating the LBP histograms at three orthogonal slices (in the case of brain images, we select axial, coronal and sagittal planes) and concatenating them. The resulting feature vector is then only three times larger than the 2D feature vector.

3.2.2   Multiscale and multimodal LBP-TOP descriptor

To incorporate multiscale information into the texture descriptor, we determine the LBP histograms at increasing radii. Examples of LBP label images at three different radii are shown in Figure 3.2.
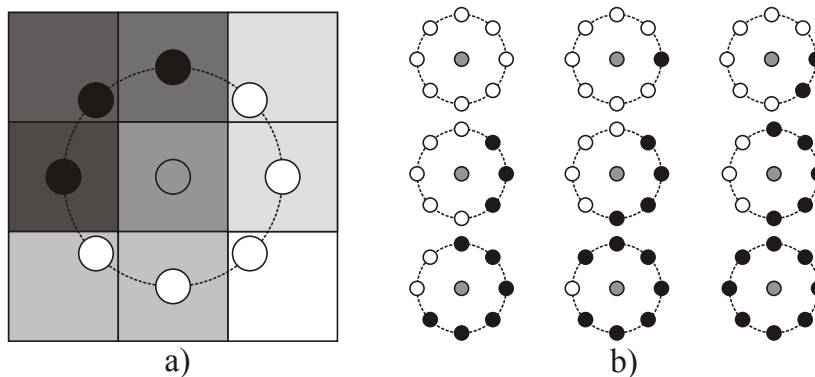
Figure 3.1: a) Local Binary Pattern neighborhood, $P = 8$, $r = 1$. The 8 circular neighbors are thresholded with the gray value of the center pixel; a binary is made by assigning zeros (black circles) to all values lower than the center pixel and ones (white circles) to those greater than or equal to the center pixel intensity; the resulting label is the decimal number that corresponds to the obtained binary; b) Rotation-invariant uniform LBPs.
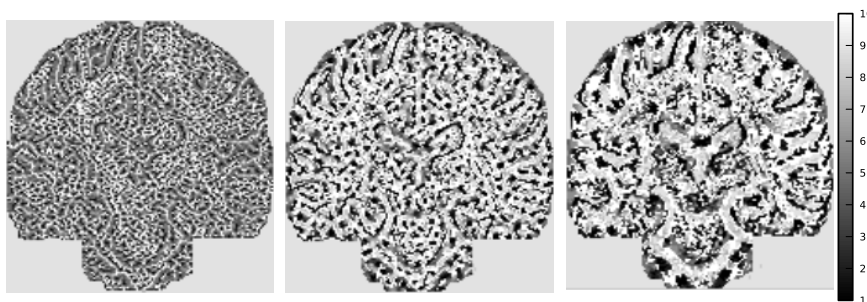

Figure 3.2: LBP labels at a coronal slice of a subject's FLAIR image, for three different radii (from left to right, $r = 1, 3, 6$).

The LBP histograms are then computed, for each subject, at the respective lesion voxels. Additionally, we include information from the three MR modalities. We concatenate all feature descriptors determined for each modality and each radius. Ultimately, for an LBP descriptor with $P = 8$ neighbors, for T modalities and R radii, the final descriptor has dimensions $3 \times (P + 2) \times T \times R$.

3.2.3 Classification

A Support Vector Machine (SVM) is a supervised classification method. Given a training set with M samples $(\mathbf{x_i}, y_i)$, $i = 1, ..., M$, with $\mathbf{x_i} \in \mathbb{R}^N$ being the training samples consisting of N features and $y_i \in \{-1, +1\}$ the respective labels, SVM finds the optimal hyperplane, $\mathbf{w} \cdot \mathbf{x} + b = 0$, which maximizes the margin between classes. For non-linear SVM, the training samples are first mapped into a higher-dimensional space by a function $\Phi(\mathbf{x})$. The two hyperplane parameters $\mathbf{w}$ and b are determined by solving a constrained minimization problem using Lagrange multipliers α_i. The final decision function is then given by [112]:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{M} y_i \alpha_i K(\mathbf{x_i}, \mathbf{x}) + b \right) \tag{3.3}$$

with the kernel $K(\mathbf{x_i}, \mathbf{x}) = \Phi(\mathbf{x_i}) \cdot \Phi(\mathbf{x})$.

3.3 Experiments and Results

3.3.1 Data

Twenty-nine datasets are retrieved from a large database of a cognition study with MCI and control subjects carried out at the University Hospital of Essen, Germany. From these 29 subjects, 15 correspond to normal controls (age 75.1 ± 3.4) and 14 to amnestic MCI subjects (age 73.4 ± 5.1).

For our analyses, we use three-dimensional isotropic images from three MR modalities (1.5 T Siemens Avanto, Germany): T1-weighted images (TR = 40ms; TE = 5ms; voxel size = 1mm^3), T2-weighted images (TR = 3200ms; TE = 416ms; voxel size = 1mm^3) and FLAIR images (TR = 6000ms; TE = 308ms; TI = 2200ms; voxel size = 1mm^3).

To extract the lesion locations in the MR images, we use the manual segmentation performed on all FLAIR images by an experienced neuroradiologist using 3D Slicer (`www.slicer.org`).

3.3.2 Preprocessing

Before any further analysis, the images must be preprocessed to eliminate variations due to the different MR acquisitions. The preprocessing steps are summarized in Figure 3.3 for the T1 image of one subject.
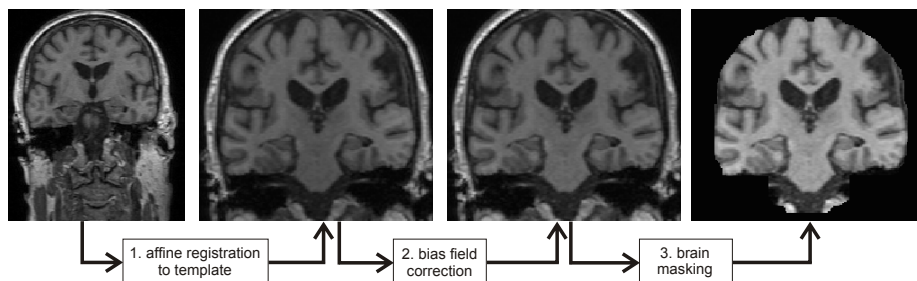
Figure 3.3: Preprocessing steps.

Firstly, we affine-register the T1 image to the Montreal Neurological Institute (MNI) Average Brain. This template corresponds to the average image of 305 T1 scans, which have been linearly transformed to the Talairach space [113]. The T2 and the FLAIR images are then affinely registered to the template-aligned T1 image. Subsequently, we correct for bias field inhomogeneities in all three images using the N4ITK method [114]. Finally, we apply the template's brain mask to the resulting images.

3.3.3 Local Binary Patterns

We start by extracting the individual LBP histograms, determined separately for each modality and neighborhood radius r, before concatenating them into a single feature vector. It is worth pointing out that we analyze only small-scale (small r values) textures, in order to ensure we are capturing only lesion-related information. Larger scales would imply that other brain structures, such as the ventricles, would also be taken into account and possibly overestimate the classifiers' performances. This is true particularly for smaller (a few voxels wide) lesions.

We use an SVM (implemented in the Python package *scikits-learn* [115]) to classify the datasets into one of the two classes: NC or MCI. To better evaluate the classifier's generalizability, we perform leave-one-out cross-validation. At each fold of the cross-validation, we scale the features such that their mean and variance over all training samples are 0 and 1, respectively. We then perform a grid search (with 5-fold cross-validation on the training set) for the best SVM parameters: kernel type - linear or Radial Basis Function (RBF); the cost C and, for the RBF kernel, the scale γ. The best classifier is then evaluated on the test

sample. The final performance measures (accuracy, sensitivity and specificity) are computed at the end of the cross-validation.

To simplify our subsequent graphical analyses, we use the balanced accuracy, which is defined as the average between the sensitivity and the specificity. The balanced accuracy is equal to the classification accuracy (rate of correctly classified subjects) when the classes are perfectly balanced (equal number of subjects in the two classes).

Results are shown in Figure 3.4. We also include the performance of the majority vote classifier that takes into account the classification outputs of the three modalities.
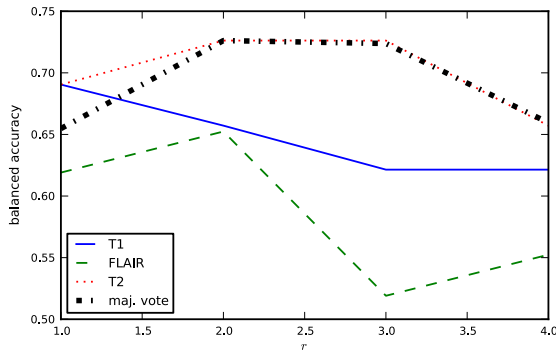
Figure 3.4: Balanced accuracy for the three MR modalities and the resulting majority vote at four LBP radii ($r = 1, 2, 3, 4$).

We observe that the accuracies obtained with the three modalities vary differently with the LBP radius. Also, the LBP descriptors extracted from the T2 images perform consistently better than the other modalities. Surprisingly, the FLAIR LBP features perform worse than those from the T1 images. The majority vote classifier does not improve the performance with respect to the individual modality classifiers, indicating that the different modalities provide the same type of information. Overall, the highest individual balanced accuracy (0.73) is obtained for the T2 images at $r = 2, 3$ (sensitivity: 0.79; specificity: 0.67).

To capture multimodal and multiscale information, we concatenate all LBP histograms into a single feature vector. For the three modalities and the four LBP radii, we obtain a 360-dimensional feature vector. We perform SVM classi-

fication on these feature vectors and obtain a sensitivity of 0.57 and a specificity of 0.67 (balanced accuracy: 0.62).

Clearly, the classification using the concatenated feature vectors has lower performance than that using the feature vectors per modality and per radius. A possible explanation is the curse of dimensionality, which occurs when the number of features is significantly larger than the number of samples. Although the SVM classifier is known to deal well with this problem, it is likely that redundant information and noise are present in the concatenated feature vector. To further analyze this effect, we perform univariate feature selection before applying the classifier.

At each fold of the cross validation, we select the k best features from the initial 360. The feature ranking is based on the F-score obtained by each individual feature on the training set. We then perform SVM classification on the reduced feature vectors. The performance of this classification is evaluated for increasing values of k (Figure 3.5).
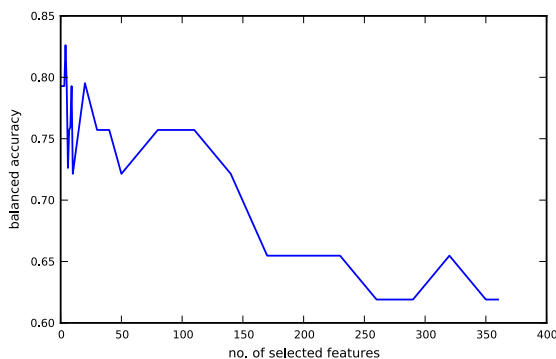

Figure 3.5: Balanced accuracy of the SVM classifier using the concatenated feature vectors after feature selection.

Performances better than those using the individual T2 LBP descriptors are obtained for fewer than 150 features, with the highest balanced accuracy (0.83, at a sensitivity of 0.79 and a specificity of 0.87) occurring at $k = 4$. As anticipated, noisy features are probably present among the entire set of 360.

We then analyze the best 4 features, selected in each cross-validation fold. We summarize this information in Table 3.1.

In one fold, a feature corresponding to $r = 1$ is also selected.

Interestingly, features from three out of the four analyzed LBP radii are con-

Table 3.1: Most frequently selected features in the cross-validation folds.

Modalities	radii	orthogonal planes	LBP labels
T2	2, 3, 4	coronal, sagittal	7, 8

sistently selected, indicating the presence of discriminative multiscale informa-
tion. Surprisingly, only the T2 images seem to contain relevant information to
distinguish between NC and MCI. Also, only the LBP histograms determined
at the coronal and the sagittal slices have bins selected. Finally, these selected
bins (corresponding to LBP labels 7 and 8) indicate that the greatest differences
between the two groups are in edge-like and more homogeneous regions. An
inspection on the LBP histograms shows that AD patients have generally fewer
of these regions, possibly indicating more heterogeneity in AD brain tissues.

3.3.4   Comparison with other methods

Volumetrics and spatial distribution

Finally, to compare the texture-based approaches with the commonly used vol-
umetric methods, we determine the total lesion load (volume) for each subject
in the two groups (Figure 3.6).

The group of NC subjects has an average lesion volume of 17.336 cm^3 ($\pm$
15.623 cm^3), while for the MCI group this average is 29.057 cm^3 ($\pm$ 50.088 cm^3).
In order to evaluate the statistical significance of this observed difference be-
tween the two groups, we perform a two-sample t-test. The obtained p-value
is 0.61, which indicates that the difference observed between the two means is
not significant.

A nearest mean classifier (in a leave-one-out cross-validation) on the WMH
volume feature yields 0.44 accuracy (0.14 sensitivity, 0.73 specificity), perform-
ing similarly to a random classifier and indicating that, for this dataset, the
lesion volume is not a discriminative feature.

Another WMH characteristic that has been investigated in the literature
is the location of the lesions in the brain. We show in Fig. 3.7 a Maximum
Intensity Projection (MIP) of the voxelwise normalized lesion count in the two
subject groups.

To quantify the significance between the voxelwise differences between the
groups, in terms of amount of lesion voxels, we perform a voxelwise t-test on
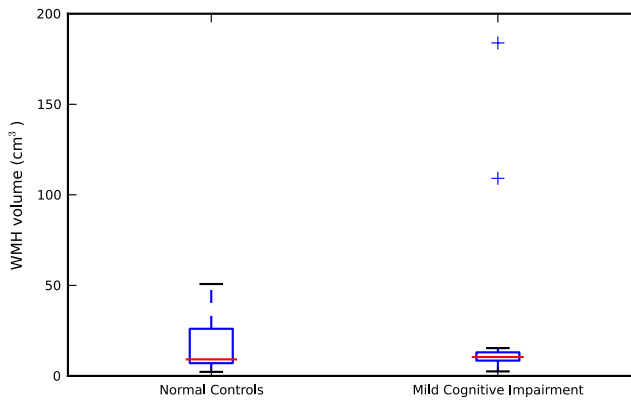
Figure 3.6: Box plots showing the total WMH volumes for all subjects: the box ranges from the lower to the upper quartile values, with a line at the median; the whiskers show the range of the data (between -1.5×IQR and +1.5×IQR; IQR: interquartile range); the crosses represent extreme data points, located beyond the whiskers' range.
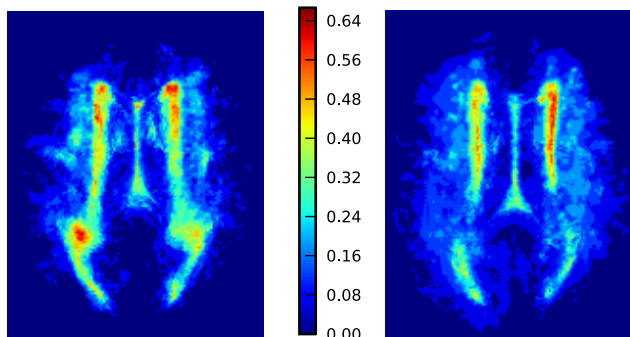

Figure 3.7: Transversal Maximum Intensity Projection (MIP) of the normalized number of lesions in each subject group. Left: normal controls; right: MCI.

the lesion spatial distribution. The resulting p-scores are shown in Figure 3.8, in a transversal minimum Intensity Projection (mIP).

Statistically significant differences in the amount of lesion voxels between the two subject groups are present mostly in periventricular regions, particularly in areas posterior to the temporal horn of the ventricles. This is in line
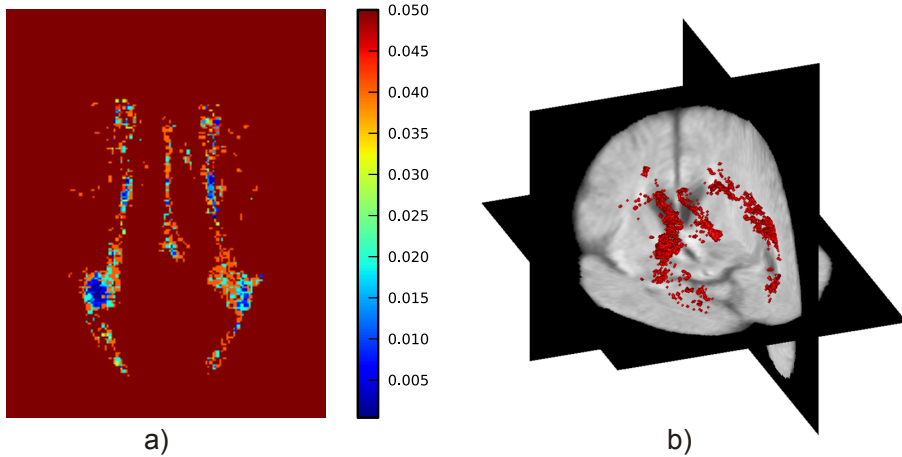
Figure 3.8: a) Transversal minimum Intensity Projection (mIP) of the p-values resulting from the statistical test between the binary GT images of the two groups, Normal Controls and Mild Cognitive Impairment; b) Isosurface at p = 0.05, displayed together with three orthogonal slices of the average FLAIR image (determined using all subjects).

with recent findings that show that a large amount of lesions in the temporal lobe is associated with the risk of developing MCI or dementia [62].

As our goal is to investigate the possibility to predict MCI based on WMH characteristics, we also use the lesion locations as features in a classification framework. The feature values are then the voxelwise binary values in the segmentation images. Similarly to what we do above, we perform univariate feature selection (in the training set, at each fold of the cross-validation) prior to SVM classification. Results are shown in Figure 3.9.

The classification based on the lesion locations is slightly better than random, reaching a maximum at $k = 1000$ (0.59, with sensitivity 0.57 and specificity 0.60). Although more informative than the simple lesion volume, the lesion locations are still far from optimal in the task of detecting MCI.

Intensity histograms

Other features that might also be informative are the actual voxel intensities. However, the MR intensities are not standardized and may vary per acquisi-
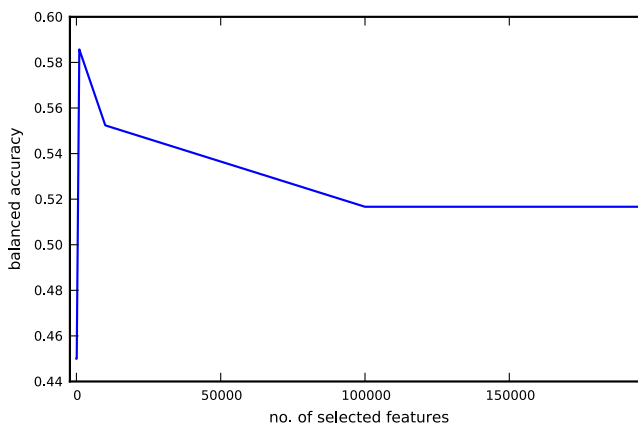
Figure 3.9: Balanced accuracy of the SVM classifier using the lesion binary values after feature selection.

tion. Therefore, a standardization step is necessary, in which the intensities are linearly transformed to a predefined range (per modality). We use the approach proposed by Nyúl *et al.* [116], in which deciles are matched to a standardized scale, determined from the average decile locations of all subjects.

Figure 3.10 shows the histograms of the images before and after standardization and Figure 3.11 shows coronal slices of the FLAIR images before and after standardization.

We then extract the histograms from the standardized images at the lesion voxels (Figure 3.12).

A question that arises when using histograms as feature vectors is the number of bins (or the bin distribution) that best describe the image's intensity content. We then evaluate the classification performance of the three MR images' histograms using different bin amounts.

We show in Figure 3.13 that, firstly, the classification rate varies significantly with the number of bins, indicating that this is a crucial parameter. Secondly, the T1 intensities perform better than the T2 and the FLAIR intensities.

Similarly to what we do with the LBP descriptors, we perform univariate feature selection before applying the SVM classifier. However, the classification rates are consistently lower than those shown in Figure 3.13 regardless of the number of selected features.

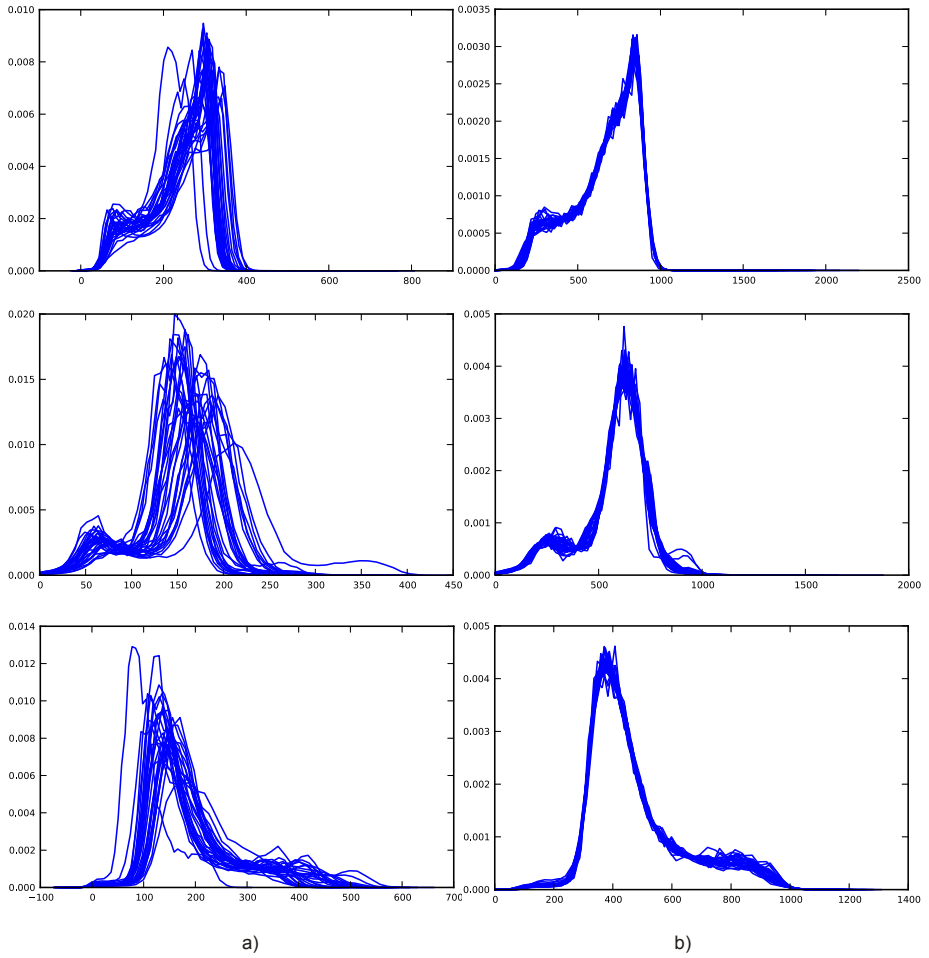A possible explanation for this performance is the way the image intensi-

Figure 3.10: Image histograms before (a) and after (b) standardization. Top: T1; middle: FLAIR; bottom: T2.

ties were standardized. Although in Figure 3.10b) the global brain histograms seem to be well aligned over all subjects, the same is not true for the local lesion histograms, shown in Figure 3.12. Particularly for the FLAIR images, we see that the lesion histograms are skewed differently for different subgroups of
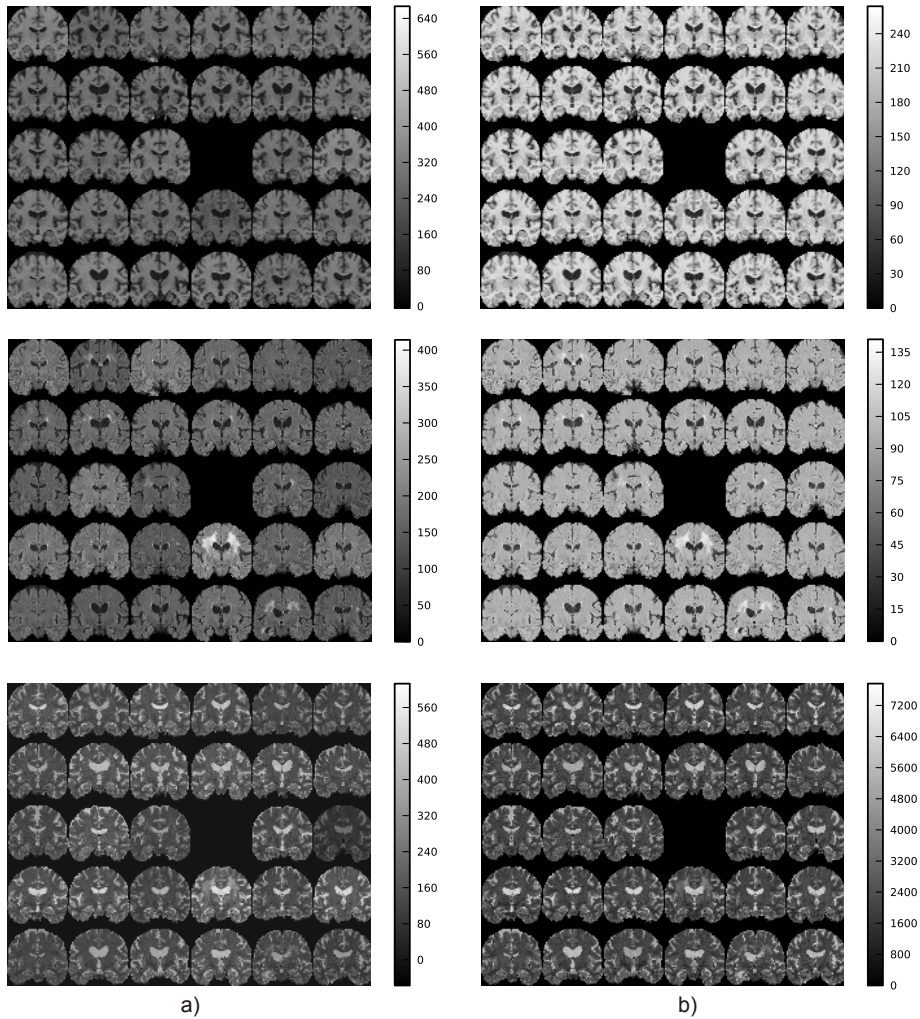
Figure 3.11: Coronal slices of all subjects: a) non-standardized; b) standardized. Top: T1; middle: FLAIR; bottom: T2.

subjects. This indicates that the widely used standardization method (Nyul's) applied here may be sub-optimal for our dataset. However, finding the right
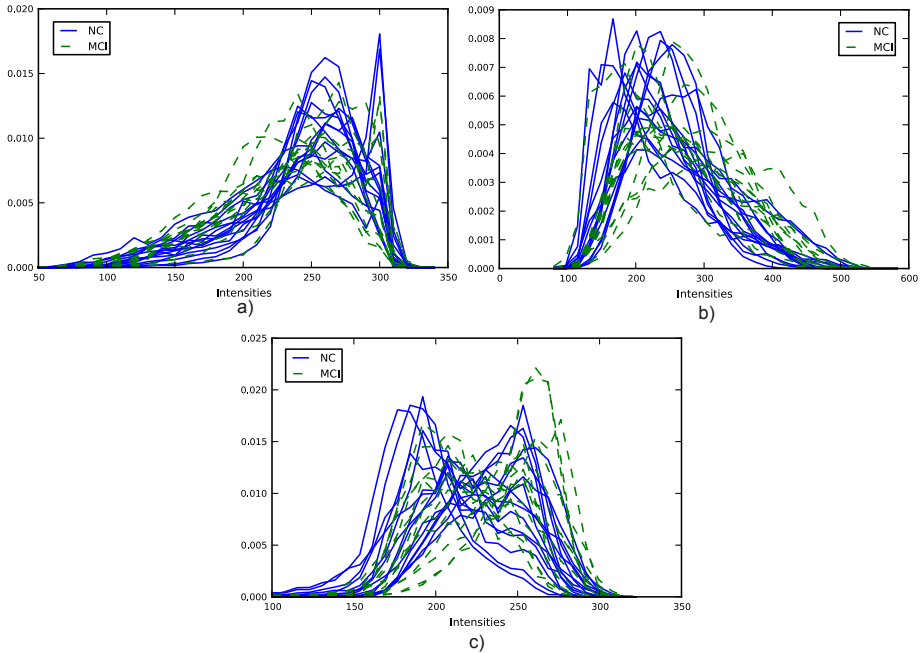
Figure 3.12: Lesion histograms from the images of each MR modality: a) T1; b) T2; c) FLAIR.

balance between a good intensity standardization and keeping the lesion intensities sensitive to the presence of disease is a complex task, particularly since no ground truth is available.

3.4   Conclusion and Recommendations

In this work we propose the use of a texture descriptor, the histogram of 3D Local Binary Pattern labels, in the detection of Mild Cognitive Impairment, based only on the brain regions containing white matter hyperintensities. We explore the incorporation of both multiscale and multimodal information, as well as the effect of feature selection on the classification performance.

We show that textural information from MR images can indeed help predict the presence of Mild Cognitive Impairment, even in a population with lesion
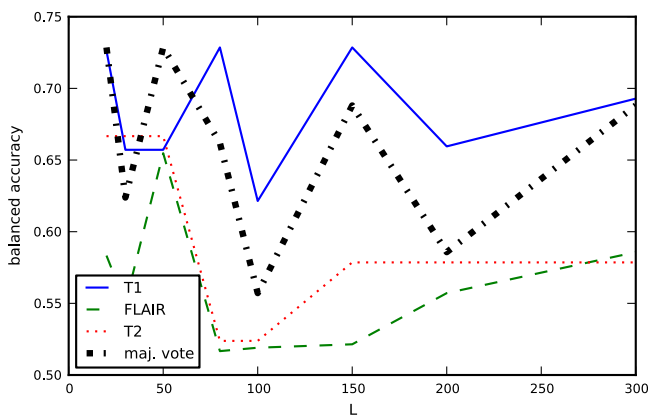
Figure 3.13: Balanced accuracy for the three MR modalities and the resulting majority vote using histograms with different numbers of bins.

loads which are not significantly different than those of a group of comparably elderly normal controls. Also, T2 images seem to provide the most discriminative features. Surprisingly, the FLAIR LBP descriptors were less discriminative than those extracted from the T1 images.

To further investigate the usefulness of including texture-based features in the quantification of WMH, studies should be performed on larger databases. In particular, images acquired at different centers should be considered to confirm the robustness of the texture descriptors to acquisition parameters.

Histogram-based features also perform better than random classification. However, such an approach requires a prior standardization step, which accounts for making corresponding tissues similar (in terms of intensity levels) across subjects. A possible consequence of this tissue matching is that disease-related differences might also be eliminated. Finding the optimal trade-off between these two effects is not straightforward. Also, a ground truth is not always available to evaluate the quality of the standardization. Additionally, we show that results are highly sensitive to the number of bins used to build the histograms. Indeed, selecting an appropriate number of bins is a complex task. For a different number of data points (in this case, lesion voxels), the number of optimal bins to represent the data in a histogram is also different. In this work, we have lesion loads ranging from only a few cm^3 to a few hundred cm^3, meaning that fixing the number of bins for all subjects is most likely sub-

optimal. Other approaches to approximate the probability density function of the lesion intensities, such as kernel density estimation, may therefore be more appropriate. Also, such descriptors are smoother than histograms, likely reducing the amount of noisy features and consequently improving classification results.

Finally, while here we focus on the analysis of the overall WMH lesions per subject, a spatial analysis of which individual lesions contribute more to the discrimination between NC and MCI is also desirable. However, that will require a larger database which includes a larger variety of lesion locations and severities in both NC and MCI subjects.

Dissimilarity-based classification using gray-level histograms

In this chapter, we begin our analysis on the use of whole image information to detect Alzheimer's disease at an early stage of development. Unlike in Chapter 3, where we have focused on the textural content of specific white matter regions (the hyperintensities), in this chapter, as well as in the following ones, we avoid making assumptions about which brain locations play an important role in the detection of early-stage Alzheimer's disease.

As an initial attempt to explore image-based features, we evaluate the performance achieved by simple gray-level histograms determined: 1) globally at the entire brain and 2) locally at patches covering the brain. In the second case, we also investigate which regions are most discriminative and whether their combination improves the final classification results. Additionally, we analyse the usefulness of a dissimilarity-based classification framework as a way of improving the feature representation of each subject. Finally, we compare our approach with a dissimilarity-based method proposed in the literature that uses a deformation-based distance measure.

The contents of this chapter have been accepted for publication at "SPIE - Medical Imaging: Computer-Aided Diagnosis 2014, San Diego, California, USA".

4.1   Abstract

Classification methods have been proposed to detect early-stage Alzheimer's disease using Magnetic Resonance images. In particular, dissimilarity-based classification has been applied using a deformation-based distance measure. However, such approach is not only computationally expensive but it also con-

siders large-scale alterations in the brain only. In this work, we propose the use of image histogram distance measures, determined both globally and locally, to detect very mild to mild Alzheimer's disease. Using an ensemble of local patches over the entire brain, we obtain an accuracy of 84% (sensitivity 80% and specificity 88%).

4.2   Introduction

Alzheimer's Disease (AD) is the most common type of dementia and a major cause of disability worldwide [117]. Its exact etiology is not yet fully understood, with multiple processes thought to be involved in the disease development. Since the initial description of dementia by Alzheimer in 1907 [118], the pathogenic mechanism that has been traditionally considered is that AD starts with the accumulation of β-amyloid (Aβ) proteins in the brain. These trigger the formation of senile plaques and neurofibrillary tangles, which in turn progressively lead to damage of the neural tissue [7]. However, recent studies show that it might be neurodegeneration that leads to the accumulation of abnormal proteins and not the other way around [6].

Early detection of AD is essential to provide the patients with adequate and timely treatments and to help monitoring their effectiveness [9]. Structural Magnetic Resonance Imaging (MRI) is a diagnostic tool that provides high-resolution images and a high brain tissue contrast. In addition, its non-invasiveness makes it a suitable imaging technique for follow-up studies.

Several pattern recognition-based methods have been developed to detect Alzheimer's disease using structural MR images. A recent comparison study of various classification methods [25] shows that the current major challenge is to discriminate patients who are at a very early stage of AD. In particular, feature extraction is a critical step in the classification framework.

Medial-temporal atrophy has been considered a valid MRI biomarker in the classification of AD [9]. Methods have been proposed that consider the hippocampal volume [27] and/or shape [28]. The volumes of the entorhinal cortex [29] and the amygdala [30] have also been used as features in AD classification. However, such volumetric approaches rely mostly on manual or semi-automatic segmentations, which are time consuming and prone to errors and inter/intra-rater variability. Also, recent metabolic studies have shown that other brain regions, such as the parietal cortex, might be more involved in the earlier stages of AD [119].

To avoid requiring a priori knowledge and segmentation of the structures

involved in the initial stages of the disease, Klein *et al.* [120] use the whole brain morphology in the classification of early-stage Alzheimer's patients. Additionally, they follow a dissimilarity-based classification approach. Dissimilarity-based classification (DBC) is a promising machine learning technique that consists of representing objects based on their distance to a representative set of objects, rather than on a set of pre-defined features [121].

In [120], the authors use a distance measure based on the whole-brain deformation field between images - the standard deviation of the logarithm of the Jacobian of the deformation field that results from non-linearly registering the two subjects' brain images.

The DBC approach with a deformation-based distance measure requires then N^2 pairwise non-linear registrations, with N being the number of subjects. For large-scale datasets, such as those available in most public databases, N is in the order of hundreds, meaning that tens of thousands of non-linear registrations need to be performed.

Another DBC application in brain MR images was proposed in [122] to detect schizophrenia, where histograms at specific regions of interest (ROIs) were used together with respective shape features. The ROIs used consisted of brain structures that had been previously manually segmented from the images by an expert.

In this work, we analyze how brain gray-level histograms perform in the detection of very mild to mild AD. We hypothesize that large-scale structural differences, such as those captured by deformation-based methods, are also reflected in terms of image intensity content. Additionally, we also use a DBC approach and compare it with the traditional classification approach. Finally, we perform local analyses at patches over the entire brain without assuming which brain structures are most discriminative.

4.3 Methods

4.3.1 Intensity standardization

MR image intensities are not standardized, meaning that the same brain tissue of the same subject in different acquisitions may have different intensity values. Since in this work we focus on the use of gray level histograms, a preprocessing step is necessary in which we transform the intensities of all subjects' images into comparable ranges.

We evaluate two histogram landmark-matching approaches: the one proposed by Nyúl *et al.* [116], in which the histogram deciles are used as landmarks, and a peak-based approach, in which the landmarks are the histogram peaks. These correspond, in turn, to the three brain tissues: cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM).

A schematic representation of the intensity transformations performed in each of the two methods is shown in Figure 4.1.
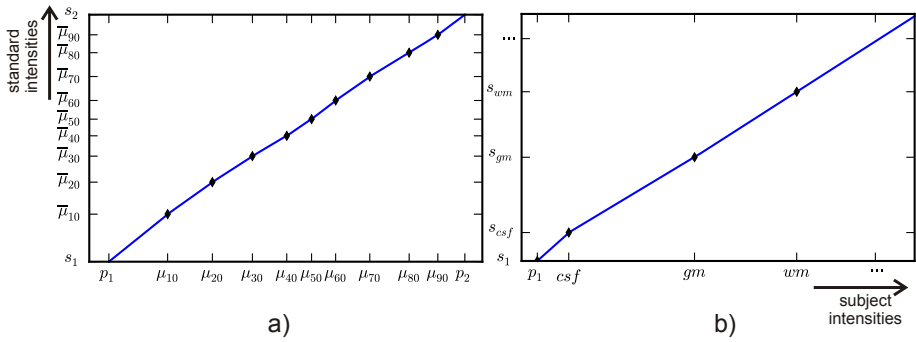

Figure 4.1: Piecewise linear intensity transformations applied to one subject's image. a) Nyúl's method: p_1 and p_2 are the lower and upper cut-off percentiles, respectively; s_1 and s_2 are the corresponding standard intensity values; $\mu_{10,\dots,90}$ are the intensities corresponding to the decile percentages; $\overline{\mu}_{10,\dots,90}$ are the deciles averaged over all subjects. b) Peak-based approach: p_1 and s_1 are the subject's and the standardized cut-off lower percentile, respectively; similarly, x and s_x are the intensities at the peaks of the subject's histogram (with x being csf, gm and wm) and the corresponding values in the defined standardized scale, respectively.

4.3.2   Distance measures

We evaluate three distance measures, defined for two normalized histograms, p and q, each determined at L bin values.

- Chi square

$$d(p, q) = \frac{1}{2} \sum_{i=1}^{L} \frac{(p_i - q_i)^2}{p_i + q_i} \tag{4.1}$$

- Euclidean

$$d(p, q) = \sqrt{\sum_{i=1}^{L} (p_i - q_i)^2} \tag{4.2}$$

- Jensen-Shannon divergence

$$d(p, q) = \frac{1}{2} \sum_{i=1}^{L} p_i \log_2 \frac{p_i}{q_i} + \frac{1}{2} \sum_{i=1}^{L+1} q_i \log_2 \frac{q_i}{p_i} \tag{4.3}$$

4.3.3 Classification in the dissimilarity space

For N subjects, we build a $N \times N$ dissimilarity matrix where each entry (i, j) corresponds to the distance between the histograms of subjects i and j. Each row (or column, since the matrix is symmetric) corresponds to the distance of the respective subject to all other subjects in the dataset.

The feature vectors are then the rows of the dissimilarity matrix [121]. We use a k-Nearest Neighbors (kNN) classifier with leave-one-out cross validation. At each fold, we estimate the optimal number of neighbors, k, in the training set.

Finally, we determine the resulting sensitivity and specificity. For most of the subsequent graphical analyses, we define the "balanced accuracy" (bac) as the average between the sensitivity and the specificity. In the case of equal-sized classes, this performance parameter corresponds to the classification accuracy.

4.4 Experiments and Results

4.4.1 Data

We use data from the publicly available OASIS (Open Access Series of Imaging Studies) database (`http://www.oasis-brains.org`). The data consist of T1 images that have been previously aligned to a template, corrected for bias field inhomogeneities and brain masked [123].

From the 436 subjects available, we use data from subjects aged 60-80 years old, divided into two classes: 66 normal controls (NC) and 70 subjects with very mild to mild AD. The designation of very mild and mild is based on the Clinical Dementia Rate (CDR). A CDR of 0 corresponds to the normal cognitive

state, 0.5 to very mild dementia, 1 to mild dementia, 2 to moderate dementia and 3 to severe dementia [124]. In this database, all subjects with dementia (CDR > 0) had been diagnosed with probable AD.

The demographic information of these subjects is summarized in Table 4.1.

Table 4.1: Demographic information of the subjects in the two classification classes.

	normal controls	early-stage Alzheimer's
Age	70.8 ± 5.5	73.2 ± 4.8
Gender (F/M)	48/18	41/29
CDR* (0/0.5/1)	66/0/0	0/48/22
MMSE**	29.1 ± 1.1	24.5 ± 4.2

*CDR: Clinical Dementia Rate
**MMSE: Mini-mental State Examination

4.4.2   Intensity standardization

The histograms of the raw and standardized images are shown in Figure 4.2.
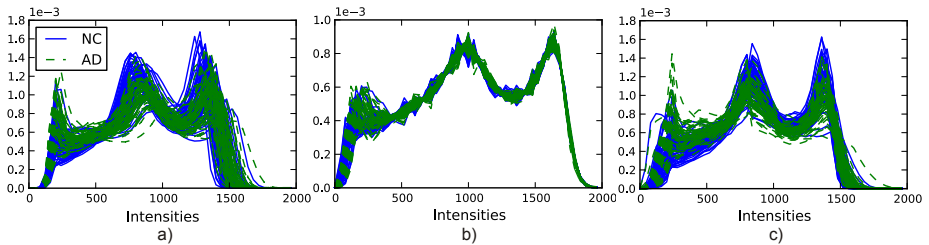

Figure 4.2: a) Raw histograms. Standardized histograms with two methods: b) Nyúl's, c) Peak-based (the left-most peak corresponds to cerebrospinal fluid, the middle peak to gray matter and the right-most to white matter). NC: normal controls; AD: Alzheimer's disease patients.

Clearly, the two standardization methods give different results. In particular, Nyúl's approach results in a closer histogram matching over the subjects, while the peak-based standardization yields more loosely matched histograms. This is to be expected considering the number of matching points used in each

of the two methods - 9 deciles in Nyúl's approach, 3 peaks in the peak-based approach. While on the one hand a closer histogram matching might indicate a more accurate standardization, it may also come with the cost of discarding relevant disease-related differences between the subjects. The effect of this step on the classification needs then to be evaluated.

In what concerns the actual histograms, a difference also seems to be present between the two groups, in both the unstandardized and the peak-based standardized images. In particular, the peak heights indicate a difference in the relative amounts of the three tissues: more CSF and less gray and white matter in AD subjects when compared to the NC subjects.

4.4.3 Dissimilarity matrix

We first make a qualitative analysis on the dissimilarity matrices obtained using the image histograms both before and after standardization (Figure 4.3). We use $L = 50$ bins and the chi square distance measure.
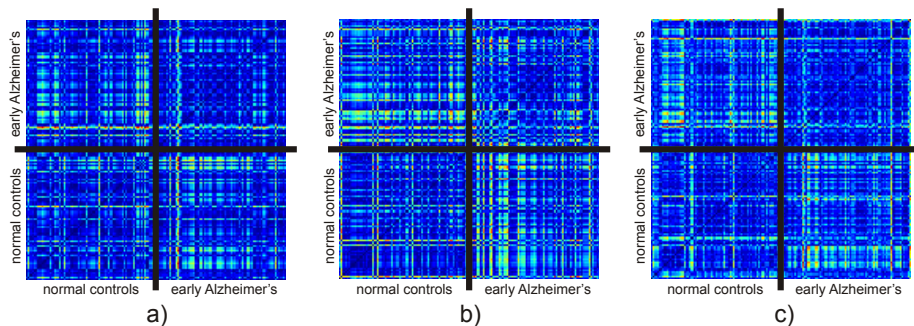

Figure 4.3: Dissimilarity matrices using a) unstandardized images and standardized images with two methods: b) Nyúl's and c) peak-based. All dissimilarity values are normalized to the maximum in the respective dissimilarity matrix to facilitate the visual comparison.

In the three cases, the dissimilarity matrices have a blockwise appearance, with the two off-diagonal blocks showing higher distances than the diagonal blocks. However, this difference is more prominent when the standardized images are used (in Figures 4.3b) and c)), which indicates that a DBC approach might perform better in these cases.

The N-dimensional rows of the dissimilarity matrix, with N being the number of subjects, are then taken as feature vectors in a leave-one-out kNN classification scheme.

4.4.4   Classification

Number of bins in the histogram

Firstly, we assess the influence of the number of bins, L, used to build the whole-brain histograms, on the classification results. To do so, we select a range of L values such that the resulting histograms are smooth and do not change significantly with L.

We perform this separately for the two standardization methods (Figures 4.4a) and b)) and compare the performance of the three distance measures described above. Also, we include the results obtained by a traditional feature-based classification, in which the histograms are directly used as feature vectors in the leave-one-out kNN classification.
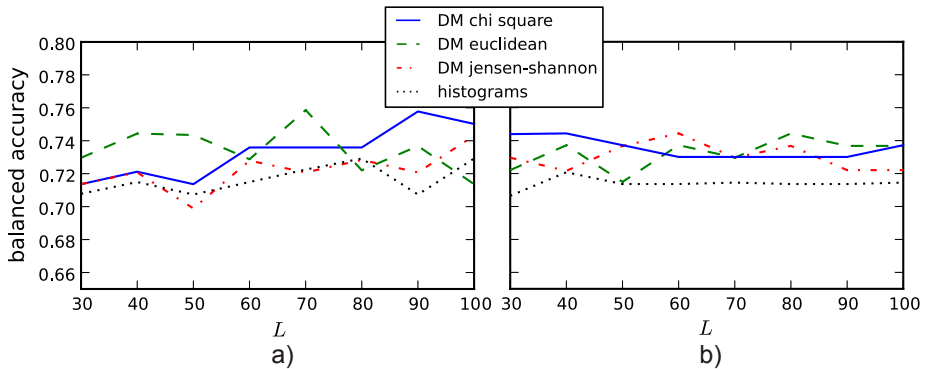
Figure 4.4: Classification rates for the DBC approach, in which the dissimilarity matrices (DM) are determined using the three distance measures, and the traditional classification using the histograms as features. The evaluation is performed at various histogram bin numbers L. a) Nyúl's, b) Peak-based.

In Table 4.2 we additionally show a summary of the results presented in Figure 4.4, computed for all L values. We also include the results obtained when using the raw intensities.

Table 4.2: Balanced accuracies for the two standardization methods, determined over all L values; μ: average, std: standard deviation, CV: coefficient of variation (std/μ).

Methods	Chi square			Euclidean			Jensen-Shannon			Histograms		
	μ	std	CV	μ	std	CV	μ	std	CV	μ	std	CV
Nyúl	0.73	0.02	2.1%	0.73	0.01	1.8%	0.72	0.01	1.6%	0.72	0.009	1.2%
Peaks	0.74	0.006	0.8%	0.73	0.009	1.2%	0.73	0.008	1.1%	0.71	0.004	0.6%
-	0.62	0.01	1.6%	0.68	0.01	2.0%	0.65	0.01	2.0%	0.71	0.01	1.2%

Interestingly, we observe that the traditional approach (histograms as feature vectors) performs similarly for both unstandardized and standardized images. This can be explained by the fact that, as shown in Figure 4.1a), the images are almost standardized. This would not be the case if they had been acquired at different imaging centers.

However, a deterioration in performance is observed when the unstandardized gray-level histograms are used in a DBC framework, indicating that such an approach requires a matching between subject intensities. Additionally, the slightly better performance obtained when using the peak-based standardization method suggests that a "looser" histogram matching may be desirable, so as to better preserve disease-related differences.

The three distance measures also perform comparably, when considering the entire range of L values. Furthermore, the variability of the classification accuracy with this parameter is always lower than 2.1%, indicating that, at least for this dataset, the number of bins used to build the histograms is not a crucial parameter in the considered range of bins.

Finally, we show in Figure 4.5 the variation of the sensitivity and specificity with the number of bins, for the two standardization methods using the chi square distance measure. We observe that, although the accuracies are similar for the two standardization methods, the peak-based standardized images yield higher specificity and lower sensitivity than those standardized using Nyúl's method, suggesting a difference in performance when standardizing NC and AD subject images.

Local patches

While the results presented above are obtained for whole-brain dissimilarities, we now investigate on local dissimilarities. These are calculated between histograms extracted from a number of cubic patches covering the entire brain.

We analyse the influence of the patch size and show the results obtained by both the best individual patch and by a combination of the patches that have the highest individual performances (Figure 4.6a)). Instead of the commonly used majority vote classifier, which considers all classifiers (in this case, all patches) with the same weight, we first apply a threshold to the individual patch accuracy in order to select only informative patches. In particular, we take the majority vote of all patches whose balanced accuracy is greater than 0.70.

In this experiment, we fix the number of bins, $L = 50$, and we use the chi
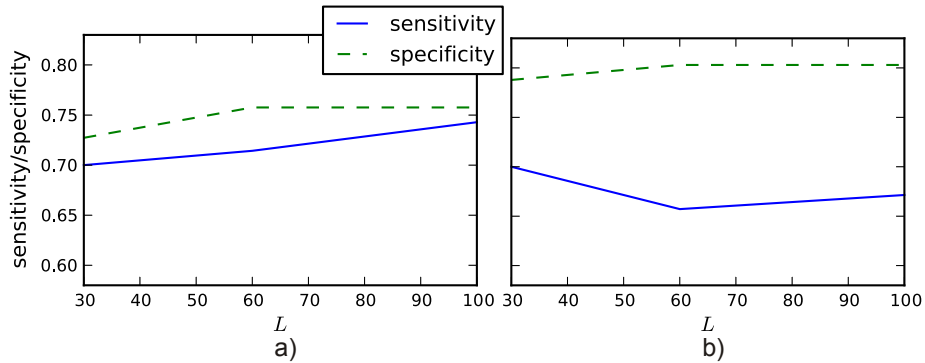
Figure 4.5: Sensitivity and specificity of the DBC approach using the chi square distance measure. The evaluation is performed at various histogram bin numbers L. a) Nyúl's, b) Peak-based.

square distance measure on the peak-based standardized images. Also, we select patches that are at least one third occupied by brain voxels.
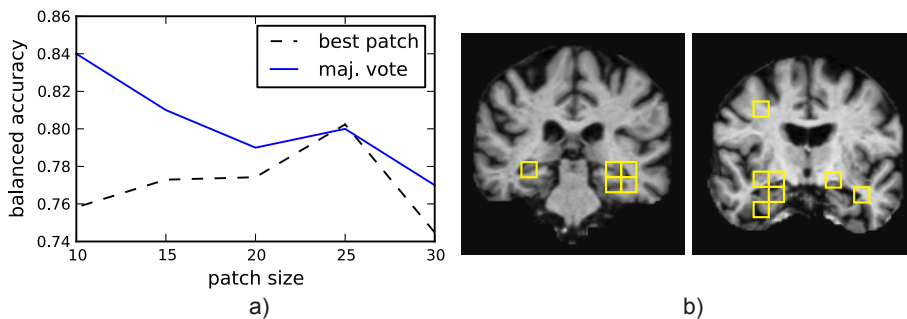
Figure 4.6: a) Balanced accuracies for the local DBC approach. b) Coronal slices of one subject highlighting the location of some of the "best" individual patches for the smallest patch size ($10{\times}10{\times}10$).

We observe that an improvement with respect to the accuracy achieved by the best patch is obtained when an ensemble of well performing patches is considered. In particular, for $10{\times}10{\times}10$ patches we obtain a balanced accuracy of 84% (sensitivity 80% and specificity 88%). The corresponding best performing

patch achieved an accuracy of 76% (sensitivity 73% and specificity 79%). These results indicate that it is useful to take different brain regions into account, since they provide complementary and discriminative information.

Additionally, Figure 4.7 shows the spatial distribution of the local patches and their respective individual balanced accuracies.
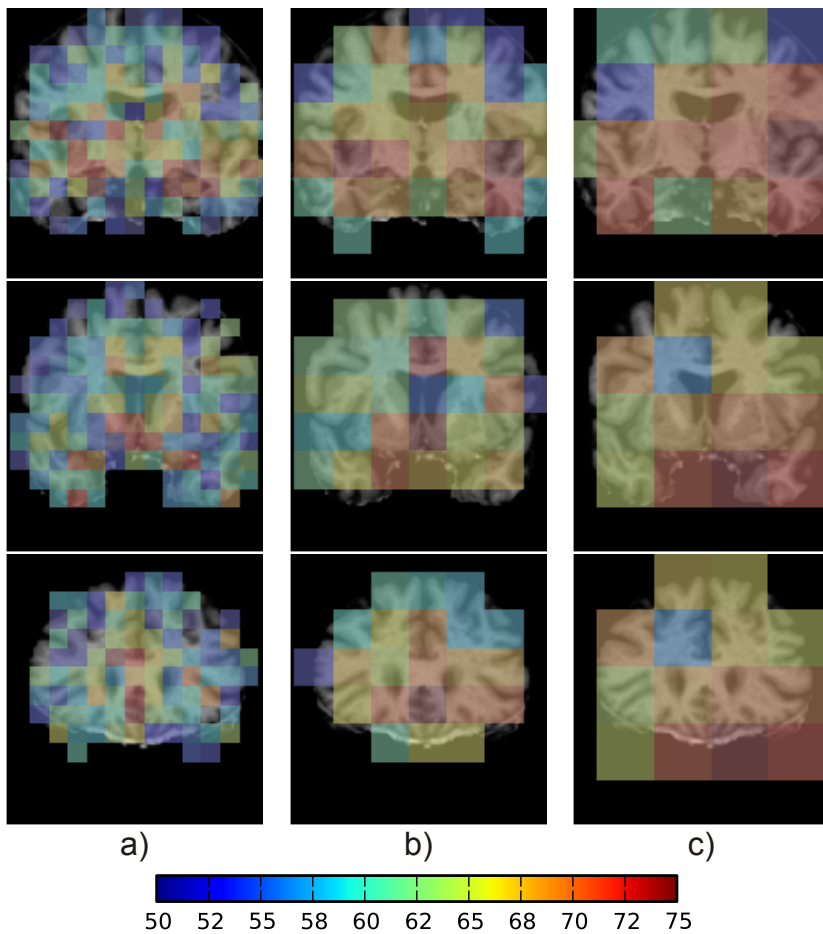

Figure 4.7: Coronal slices overlayed by the accuracy maps, determined at local patches of different sizes: a) 10×10×10, b) 20×20×20, c) 30×30×30.

The "best" patches are located near the medial temporal lobe, particularly in the hippocampal region. This is more clear when we consider small patches (Figure 4.7a)), which allow for a higher resolution of the accuracy maps. For larger patches, also the ventricular and cortical regions show as highly discriminative. This is possibly due to the different proportions of CSF and GM inside those patches, caused by more expanded ventricles in the early-stage Alzheimer's subject group.

Comparison with other methods

To compare our approach with that proposed by Klein *et al.*, we apply Klein's method to a subset of the dataset we use above. For practical reasons, we select 40 NC and 40 AD subjects. As mentioned above, this approach is particularly time-consuming, since it requires a large number of non-linear registrations (N^2, with N being the number of subjects).

We perform B-spline registration using the same Elastix [125] parameter file as that used in [120] (available at `http://elastix.bigr.nl/wiki/index.php/Par0010`).

The distance measure is then determined as follows:

$$d(i,j) = \frac{1}{2}(\sigma(\log J_{ij}) + \sigma(\log J_{ji})) \tag{4.4}$$

with J_{ij} being the Jacobian determinant of the transformation field that results from warping image i to image j; σ is the standard deviation.

We obtain a sensitivity of 67.5% and a specificity of 65% (balanced accuracy of 66.3%). By applying the global approach we have proposed to the exact same subjects, the resulting sensitivity and specificity are 75% and 72%, respectively (balanced accuracy of 73.8%).

4.5   Conclusion and recommendations

In this work, we investigate the use of simple gray-level histograms to detect Alzheimer's disease at an early stage of development. We propose two approaches for determining the histograms: a whole-brain approach and a local patchwise approach. We show that both perform reasonably well in this classification task, and that local patches can be combined to improve the classification. Also, the dissimilarity representation performs better than the individual

histograms used as feature vectors, corroborating findings from other studies on brain MR images.

However, a major drawback of histogram-based approaches is that the image intensities need to be standardized, so that there is a meaningful correspondence across all subjects. Finding the most appropriate standardization method is a difficult problem enhanced by the fact that often no ground truth exists.

Also, selecting an appropriate number of histogram bins is not straightforward. The optimal amount of bins is likely to vary with, for example, the patch size - the smaller the patch, the fewer the data points and, consequently, the fewer the bins desirable to build the histograms. We show that for our dataset, however, the variability of the classification accuracy with the number of bins is always lower than or about 2%. Future studies are necessary to confirm the robustness of the proposed method with this parameter.

Additionally, future work will include exploring other image-based (preferably robust to the intensity scale) descriptors such as texture features. A more precise localization map can be achieved with smaller patches, at the cost, however, of a lower robustness of the histogram representation. Other patch combinations, such as overlapping patches and/or patches of different sizes, can therefore overcome this problem. Also, in this work we apply a fixed threshold to select only the best patches for the final classifier. Other combination rules, such as adaptive weighting, should be further explored.

Second-order statistical texture maps

In Chapter 4, we have shown that image intensities are informative and that their histograms can be used as features in the classification of early-stage Alzheimer's disease. This chapter further introduces the use of second-order statistical texture features that take pairwise gray-level occurrence frequencies into account. We determine these textures within local patches centered at each brain voxel to obtain the texture maps. Similarly to what is done in morphometric studies, we also non-linearly register the texture maps into a common template and compare their performance with that achieved by the gray matter density map used in voxel-brain morphometry approaches. With this, we aim to compare the discriminative power of two voxelwise measures - the local gray matter concentration and local statistical textures - after performing the same exact preprocessing steps.

This chapter is based on the following publication: *Lopes Simoes, A.R. and Slump, C.H. and van Cappellen van Walsum, A.-M. (2012) Using local texture maps of brain MR images to detect Mild Cognitive Impairment. In: 21st International Conference on Pattern Recognition (ICPR 2012), 11-15 Nov 2012, Tsukuba, Japan. pp. 153-156.*

5.1 Abstract

Early detection of Alzheimer's disease is expected to aid in the development and monitoring of more effective treatments. Classification methods have been proposed to distinguish Alzheimer's patients from normal controls using Magnetic Resonance Images. However, their performance drops when classifying patients at a prodromal stage, such as in Mild Cognitive Impairment. Most often, the features used in these classification tasks are related to structural measures such as volume, shape and tissue density. However, microstructural

changes have been shown to arise even earlier than these larger-scale alterations. Taking this into account, we propose the use of local statistical texture maps that make no assumptions regarding the location of the affected brain regions. Each voxel contains texture information from its local neighborhood and is used as a feature in the classification of normal controls and Mild Cognitive Impairment patients. The proposed approach obtained an accuracy of 87% (sensitivity 85%, specificity 95%) with Support Vector Machines, outperforming the 63% achieved by the local gray matter density feature.

5.2 Introduction

Alzheimer's Disease (AD) is the most common type of dementia and a major cause of disability worldwide [117]. Early detection of AD is essential to provide the patients with adequate and timely treatments and to help researchers monitor their effectiveness. Structural Magnetic Resonance Imaging (MRI) is a diagnostic tool that provides high-resolution images and a high brain tissue contrast. In addition, its non-invasiveness makes it a suitable imaging technique for follow-up studies.

A limitation of most state-of the-art MR image analysis methods in this field is that they often concern only group comparisons. Although these methods can provide a description of the location and magnitude of statistically significant differences between two groups, they have limited clinical value for individual patients.

This limitation has led to the development of classification methods to identify Alzheimer's patients from Normal Controls (NC) and, more recently, to distinguish NC from patients suffering from Mild Cognitive Impairment (MCI), which indicates high risk of developing Alzheimer's. As pointed out by a recent comparison study on various classification methods [25], the current major challenge is to discriminate patients who are at a very early stage of AD or even possibly before they start developing the disease. As shown by the comparison results, the performance of most classifiers dropped significantly when they attempted to classify between NC and MCI.

Typically, the features used by these classification methods concern the volume and/or the shape of specific brain structures, like the hippocampus [25]. Voxel-Based Morphometry (VBM) approaches have also been used, which analyze the local concentration of gray matter [25, 37].

However, such tools are limited by the segmentation quality of the structures of interest. Furthermore, it has been shown that the brain microstructure

starts to deteriorate several years before the first symptoms arise and before structural alterations can be detected [9].

Texture Analysis (TA) is an image processing tool that has recently found applications in the study of various neurological diseases, including Alzheimer's. It extracts information that is otherwise not visible by a direct analysis of the image intensity and shape properties. In [48], the authors performed 2D texture analysis using the entire brain to classify between AD and NC. Because the whole brain was used, no discrimination between significant regions was performed. In [52], Zhang *et al*. also classified patients as AD or NC using 3D texture features computed at manually defined spherical Regions of Interest (ROIs), in the hippocampus and the entorhinal cortex. However, and as the authors recognized, the results varied significantly with the location and the size of the chosen ROI. Furthermore, in neither of these two studies was an analysis with MCI patients performed. Other studies have carried out texture analysis in the corpus callosum and thalamus [53]. In all cases, the texture descriptors are computed at manually segmented ROIs, thereby requiring a priori knowledge about the disease and becoming dependent on the quality of the segmentations. Also, to the best of our knowledge, no comparisons between the two approaches (structural and textural) have been performed.

In this work, we propose the use of *local* statistical (co-occurrence matrix based) texture maps as features to be used in the classification of NC and MCI. In these maps, each voxel contains texture information from its local neighborhood and is considered as a feature for classification. We perform a statistical significance analysis on these voxels as a feature selection step. Finally, we use Support Vector Machines (SVM) in a cross-validation scheme to classify the subjects. We compare our method with a structural approach that uses, as features, the voxels in the gray matter probability map [37].

Our contributions are the following: application of local statistical texture maps to the classification of NC and MCI, which make no assumption about the expected location of significant differences and that require no previous segmentation of brain structures; performance comparison of the proposed features and a widely used structural feature - the local gray matter density. To the best of our knowledge, no other texture studies have made such comparison.

5.3   Methods

5.3.1   Calculation of the feature maps

The Haralick features are based on the Gray Level Co-occurrence Matrix (GLCM), which gives information about the statistical distribution of voxel intensity pairs [126]. In this work, we refer to these texture descriptors by the following: F1 - angular second moment; F2 - contrast; F3 - correlation; F4 - sum of squares; F5 - inverse difference moment; F6 - sum average; F7 - sum variance; F8 - sum entropy; F9 - entropy; F10 - difference variance; F11 - difference entropy. For a complete description of the features, we refer the reader to [126].

As in previous texture studies [52, 53], we compute the first eleven Haralick features (according to [126]) at a $3 \times 3 \times 3$ sliding window centered on each brain voxel. This allows for texture analysis in the entire brain rather than at specific ROIs. The GLCM is determined for all 13 three-dimensional directions, considering voxel pairs at a distance of 1 voxel. In order to increase the computational speed of these calculations, and to avoid very sparse GLCMs, we quantize the original image intensities to 5 bits (range $[0, 31]$). After texture feature calculation, we obtain, for each subject, 11 feature maps.

5.4   Experiments and Results

5.4.1   Data and preprocessing

For this study, datasets from 15 Normal Controls (75.4±4.5 years, 8 males and 7 females) and 15 MCI patients (73.3±8.2 years, 10 males and 5 females) were retrieved from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [127]. The data consist of three-dimensional T1 images acquired at 3T. These images have been previously corrected for acquisition artifacts such as bias field inhomogeneities, geometric distortions and scaling, as described in [127]. To eliminate global differences between brain shapes and volumes, we align all images to the same spatial reference using a non-linear diffeomorphic registration method, DARTEL [128].

5.4.2   Feature maps

We then register the features maps into the template space, by applying the same warp field that originated from the non-linear registration of the T1

images. An 8mm (FWHM) isotropic Gaussian kernel is finally applied to smoothen the aligned feature maps.

To obtain the gray matter density feature maps, we first segment the brain images using the probabilistic segmentation method offered by SPM8 (Wellcome Trust Centre for Neuroimaging, `http://www.fil.ion.ucl.ac.uk/spm`). Then, and similarly to what is done with the texture maps, we apply the respective warp field obtained in the non-linear registration step to the gray matter segmentations, followed by Gaussian smoothing. The resulting maps represent the local concentration of gray matter per voxel. Two-dimensional slices of all obtained feature maps are shown in Figure 5.1.
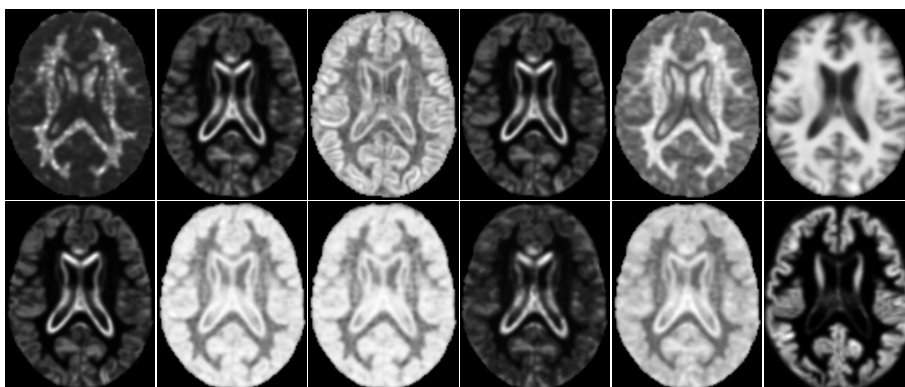

Figure 5.1: Feature maps of an MCI patient after non-linear registration to a common spatial reference. Bottom right: gray matter tissue probability map.

5.4.3 Classification

We use an SVM (implemented in the Python package *scikits-learn* [115]) to classify the datasets into one of the two classes: NC or MCI. To better evaluate the classifier's generalizability, we perform a random subsampling evaluation with 10 random permutations, in which the test set corresponds to 10% of the data samples. At each training fold, we carry out an analysis of variance (ANOVA) test on the training samples and select the 5% most significant voxels, which are then used as features in the classification task. We perform a grid search (with 5-fold cross-validation on the training set) for the best SVM parameters: kernel type - linear or Radial Basis Function (RBF); the cost C and,

for the RBF kernel, the scale γ. The best classifier is then evaluated on the test set. The final performance measures (accuracy, sensitivity and specificity) are computed as the average of the values obtained at each evaluation fold.

The classification results are shown in Figures 5.2 and 5.3. The texture descriptor with the best performance (F3 - correlation) achieved a mean accuracy (percentage of correctly classified subjects) of 87%, at a sensitivity of 85% and a specificity of 95%. In contrast, the accuracy of the structural feature was 63%, with 75% sensitivity and 55% specificity. A Wilcoxon-Mann-Whitney statistical significance test on the evaluation folds' results showed that feature F3 significantly outperformed the gray matter density feature in terms of accuracy ($p = 0.007$) and specificity ($p = 0.01$). Feature F8 (sum entropy) showed also, at a high significance level ($p = 0.06$), a better accuracy than the structural feature.
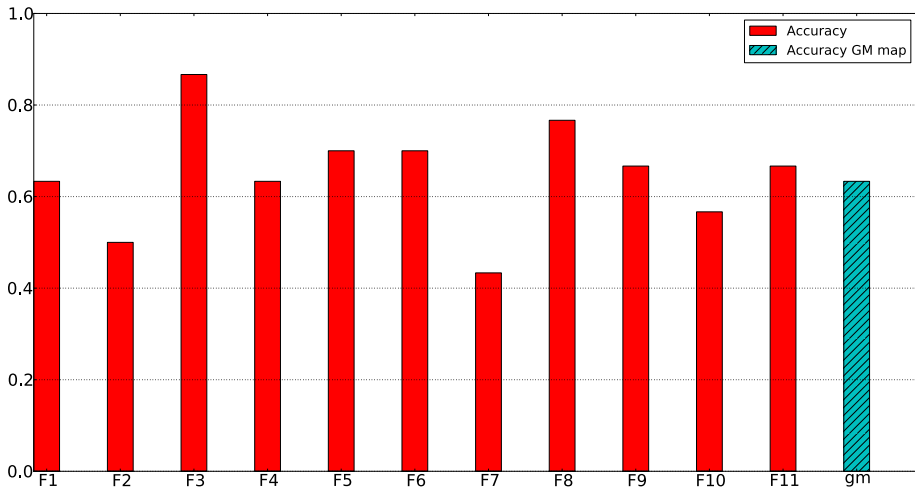
Figure 5.2: Mean accuracy obtained using the first 11 Haralick features and the Gray Matter (GM) tissue probability maps (rightmost blue hatched bar).

In addition, we show the brain voxels that were selected by the ANOVA test in one of the training folds (Figure 5.4a). We observe that using the correlation (F3) map voxels in the left hippocampus are detected as being statistically significant (and consequently used in the classification). Voxels in the brain ventricles, particularly near the edges, are also selected, as well as in the white matter and near the lateral sulcus. The higher accuracy of the classification us-
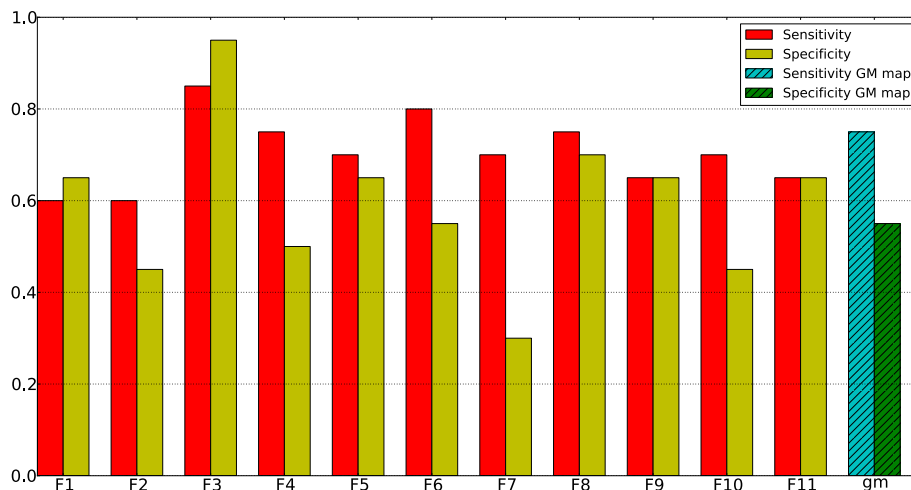
Figure 5.3: Mean sensitivitiy and specificity obtained using the first 11 Haralick features and the Gray Matter (GM) tissue density maps (rightmost green and blue hatched bars).

ing this feature map, when compared to the structural feature, indicates that these regions might play a role in MCI, even though their corresponding gray matter density is not significantly different between the groups. As a comparison, we show, in Figure 5.4b), the statistical differences between the same NC group and a group of 12 AD patients, where we clearly see, for both feature types, the two hippocampi being selected (the left being more significant). MRI signal changes which do not correlate with structural measurements have already been observed in ageing subjects [44]. The underlying cause for these alterations lies probably in the change of water, protein and mineral content of the tissues. A similar explanation can be given to why texture descriptors might be able to capture early signals of dementia.

A final analysis was performed on the effect of varying the percentile of features selected for classification, although no significant changes in classification performance were obtained.
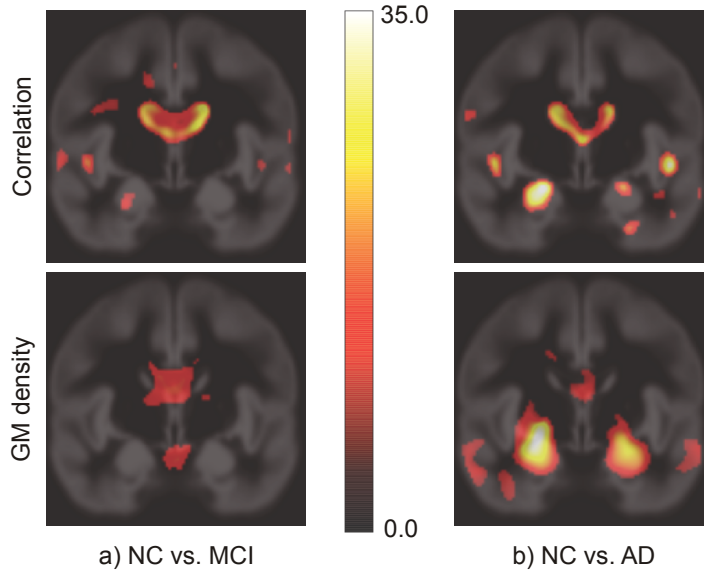
Figure 5.4: F-values of the statistical test (only the 5% most significant voxels are shown) for the correlation map and the GM local density map.

5.5   Conclusions and recommendations

In this work, we have analyzed the performance of statistical texture maps in classifying MCI patients and normal elderly controls. We used a whole-brain voxelwise approach, in which we made no assumptions about the expected location of differences between the two subject groups.

We obtained a mean accuracy of 87% (sensitivity of 85% and specificity of 95%) when using the correlation map voxels as features in an SVM classification task, outperforming the structural feature map - the local gray matter density. Remarkably, the voxels selected using the two feature maps were not the same, suggesting that texture- and structure-based features might be sensitive to distinct aspects of the disease. In particular, part of the left hippocampus was selected when using the texture map but not with the GM density map, possibly indicating an earlier sensitivity of the texture descriptor to changes in this region.

Further work will include a more thorough evaluation of other classifiers

and feature selection/extraction methods. Also, the influence of the preprocessing steps on the classification performance should be assessed. This includes the non-linear registration to the common spatial reference and the smoothing applied to the registered feature maps.

The influence of the size of the local window chosen to compute the features should be evaluated. In this work, we focused on very fine-scale statistical textures. A multi-scale analysis will provide further insight on also larger-scale texture properties. Other feature types, such as higher-order statistical features and Gabor wavelets, as well as combinations of various features, need also be considered.

Additionally, a comparison between the results obtained with images acquired at 3T and at the most commonly available 1.5T is desirable. In particular, it is worth investigating how both structure- and texture-based features perform at the two field strengths. Similarly, other MRI modalities (such as T2 images) should be considered.

Finally, the classification must be performed with a larger number of samples to allow for stronger conclusions. However, these preliminary results seem to indicate that microstructural information, such as that provided by local texture descriptors, can play a useful role towards better and earlier detection of Alzheimer's disease.

Local Binary Patterns in local patches

In this chapter, we propose another texture-based approach to detect early-stage Alzheimer's disease. While the features used in Chapters 4 and 5 depend directly on the original gray-scale values of the images, in this chapter we use intensity-invariant descriptors, the Local Binary Pattern histograms, which have been introduced in Chapter 3. Similarly to Chapter 4, but unlike what has been done in Chapter 5, we avoid non-linear registrations and use only affine-registered images. Furthermore, we explore the patchwise approach introduced in Chapter 4 and perform an exhaustive search for discriminative regions over the whole brain. Finally, we compare our results to another structural-based approach (the volumes of the hippocampi and the amygdalae) and to other methods in the literature that also use local features.

This chapter has been submitted to Neuroradiology.

6.1 Abstract

Early detection of Alzheimer's Disease is expected to aid in the development and monitoring of more effective treatments. Classification methods have been proposed to distinguish Alzheimer's patients from normal controls using Magnetic Resonance images. Most often, the features used in these classification tasks are related to structural measures such as volume, shape and tissue density. However, these approaches rely on *a priori* assumptions about the disease location and/or non-linear registrations to a template. In this work, we propose the use of 3D texture analysis using Local Binary Patterns computed at a large number of image patches and combined in a classifier ensemble. We evaluate our method in a publicly available database of 66 normal controls and 70 very mild to mild Alzheimer's patients. For an ensemble of small ($10 \times 10 \times 10$ voxels) highly overlapping patches, the classification accuracy is 100% in a

leave-one-out cross-validation setting. Finally, we show that our method is able not only to perform accurate classification but also to localize discriminative brain regions, which are in accordance with the medical literature.

6.2   Introduction

Alzheimer's Disease (AD) is the most common type of dementia and a major cause of disability worldwide [117]. Its exact etiology is not yet fully understood, with multiple processes thought to be involved in the disease development. Since the initial description of dementia by Alzheimer in 1907 [118], the pathogenic mechanism that has been traditionally considered is that AD starts with the accumulation of β-amyloid (Aβ) proteins in the brain. These trigger the formation of senile plaques and neurofibrillary tangles, which in turn progressively lead to damage of the neural tissue [7]. However, recent evidence shows that it might be neurodegeneration that leads to the accumulation of abnormal proteins and not the other way around [6].

Early detection of AD is essential to provide the patients with adequate and timely treatments and to help monitoring their effectiveness [9]. Structural Magnetic Resonance Imaging (MRI) is a diagnostic tool that provides high-resolution images and a high brain tissue contrast. In addition, its non-invasiveness makes it a suitable imaging technique for follow-up studies.

Several pattern recognition-based methods have been developed to detect Alzheimer's disease using structural MR images. A recent comparison study of various classification methods [25] shows that the current major challenge is to discriminate patients who are at a very early stage of AD. In particular, feature extraction is a critical step in the classification framework.

Medial-temporal atrophy has been considered a valid MRI biomarker in the classification of AD [9]. Methods have been proposed that consider the hippocampal volume [27] and/or shape [28]. The volumes of the entorhinal cortex [29] and the amygdala [30] have also been used as features in AD classification. However, such volumetric approaches rely mostly on manual or semi-automatic segmentations, which are time consuming and prone to errors and inter/intra-rater variability. Also, recent metabolic studies have shown that other brain regions, such as the parietal cortex, might be more involved in the earlier stages of AD [119].

To overcome the limitations of ROI-based approaches, methods that consider voxelwise measures can be employed [129]. A widely used approach is to extract the voxelwise probability of the three brain classes. This technique

is called Voxel-Based Morphometry (VBM) [38] and is generally followed by classification with Support Vector Machines (SVM). The gray matter probability map (often referred to as "density" or "concentration" map) is the most often used, based on the assumption that AD primarily affects the cortical structures, as a consequence of the underlying neuronal loss [37]. Similarly, Deformation-Based Morphometry (DBM) considers the voxelwise deformation field that results from warping each subject's image to a common template [39, 40]. With these voxel-based methods, it is possible to determine which voxels are more discriminative between the subject groups, and maps showing the brain regions that are related to the disease can be created.

However, VBM and DBM-based approaches always require non-linear alignments to a template, in order to achieve voxelwise inter-subject correspondence [36]. A drawback is that, due to the high anatomical variability of brain structures, it is difficult to evaluate the accuracy of this inter-subject matching. Furthermore, while elastic registration can give more precise alignment results than, for example, affine registration, there is also the risk of a misalignment [42] or an over-alignment which can result in informative patterns being eliminated from the images [43].

To overcome these limitations, [43] have proposed the use of local features, which are computed independently for each subject. In particular, they have used the well-known Scale Invariant Feature Transform (SIFT) features [130] to classify AD. Further adaptations of Toews' approach were proposed by [131] and [132].

Local Binary Patterns (LBPs) are a computationally efficient method for describing the local texture of an image. Since they were first proposed by [108], LBPs have been widely used in various computer vision applications such as face detection and recognition [133]. LBP descriptors are invariant to intensity changes and can also be rotation-invariant [110], which makes them an attractive alternative for texture analysis of medical images. In particular, their robustness to various artifacts that are commonly present in brain MRI has been studied by [134]. Also in neuroimaging, LBPs have recently been used as descriptors for content-based image retrieval [135]. [136] have proposed a 3D LBP variant to classify Attention-Deficit/Hyperactivity Disorder (AHDH) in brain MR images. [67] performed classification between normal controls and AD subjects based on LBP descriptors computed at the subjects' white matter hyperintensities.

Two main approaches can be followed when using LBPs as image descriptors - we can either compute them at the whole image or at specific local regions. As an analogy with the face recognition literature, we point to holistic

or local methods, with the latter performing generally better than the former [137]. In this work, we are particularly interested in investigating which brain regions are affected at the early stage of AD. Because we also want to avoid making *a priori* assumptions about these locations, we use a patchwise approach, in which LBP descriptors are determined for each local patch in the image.

A question that arises when using such patchwise features is how to combine them for the final classification. Many approaches, particularly in the face recognition literature, concatenate the feature descriptors into a single large feature vector [138]. However, when a large number of patches are defined, the classification problem becomes high dimensional, and the "curse of dimensionality" problem arises [137]. Since one of our goals is to obtain a localization map of discriminative regions, we need a fine image sampling, i.e., many patches at various locations in the brain image. Feature concatenation would therefore not be appropriate.

Another approach is to use an ensemble of patchwise classifiers, which can reduce the feature dimensionality per classification. The assumption is that the combination of complementary information can give more accurate results [139].

The classifier AdaBoost [140] has been used with this goal in face recognition applications [141]. It consists of iteratively adding weak classifiers to the ensemble and adjusting the weights attributed to the training samples, according to whether they have been mis- or correctly classified. Although such methods can alleviate the "curse of dimensionality", their training step can also be very time consuming when a large number of base classifiers (patches) is used, since for each iteration all patches must be classified.

In this work, we propose the use of a 3D variant of the LBP descriptor [111], computed at local image patches densely sampled in the entire brain. We classify each patch with a template matching classifier and combine the results by fixed weighted majority voting.

We evaluate our method on data retrieved from a publicly available database, in which images have been affine-registered to a template. We hypothesize that there is no need for non-linear registration since the image patches contain already discriminative information between the groups of elderly controls and early-stage Alzheimer's. We compare our results with those obtained in recent studies on the same database - both shape/volume-based methods and others that, like ours, use local features.

Finally, besides classifying between very mild to mild AD and normal con-

trols, we also analyze which local brain regions are most discriminative in the two subject groups.

6.3 Methods

6.3.1 Local Binary Patterns

The local binary pattern (LBP) is an operator that labels the pixels of an image according to its local appearance, or small-scale structure. The most common texture descriptor based on the LBP operator is the histogram of the labels within a region of interest [107].

The formal description of such an operator is as follows. At a pixel with gray value g_c, with P neighbors at a neighborhood radius of r, its LBP label is given by:

$$LBP_{P,r} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{6.1}$$

with

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.2}$$

In its original form, proposed by [108], the LBP label of a certain pixel is determined by thresholding its 8 neighbors (in a 3×3 neighborhood) with its own gray value and converting the resulting binary into a decimal number. More recently, several variants of the LBP descriptor have been proposed [107]. In particular, [109] extended the LBP descriptor to allow for other neighborhood radii, r, and more sampling points, P. Typical values of r are in the order of few pixels. Accordingly, P takes usually values in the order of the number of neighbors for a certain r.

Figure 6.1a) shows an example of an LBP neighborhood. The circles that, after bilinear interpolation of the respective neighborhood, have a lower intensity than the center pixel are shown in black and the others in white.

More recently, it was observed that the majority of the observed patterns are uniform (described by no more than two bitwise 0-1 transitions). For the 8-bit ($P = 8$) LBP descriptor, 58 out of the 256 possibilities correspond to uniform patterns. Furthermore, it is possible to obtain a rotation-invariant descriptor by rotating the LBP binary code into its minimum decimal value [110]. In general, for P neighbors we have $P + 1$ possible uniform rotation-invariant LBP labels

(shown in Figure 6.1b)). All non-uniform patterns are assigned label $N = P + 2$.

In this work, we are particularly concerned with the computational cost of the feature extraction step, since we perform local analysis on a large number of 3D image patches. Also, we want the feature descriptor to be as insensitive as possible to minor errors in the affine registration to the template. Therefore, we use the uniform rotation-invariant LBP descriptor. An analysis of the influence of the neighborhood parameters, P and r, on the classification accuracies is performed in Section 6.4.
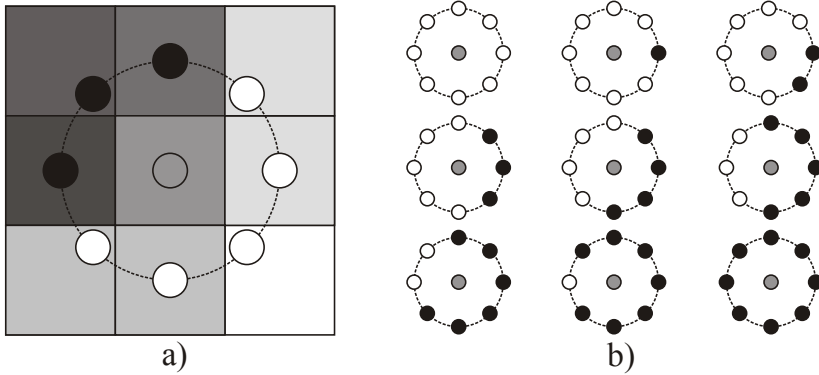

Figure 6.1: a) Local Binary Pattern neighborhood, $P = 8, r = 1$. The 8 circular neighbors are thresholded with the gray value of the center pixel; a binary is made by assigning zeros (black circles) to all values lower than the center pixel and ones (white circles) to those greater than or equal to the center pixel intensity; the resulting label is the decimal number that corresponds to the obtained binary; b) Rotation-invariant uniform LBPs.

6.3.2 LBP-TOP

Two extensions of the LBP operator to 3D data have been proposed in [111], mostly for spatio-temporal applications. In particular, LBP-TOP (Three Orthogonal Projections) is computationally simple. It consists of calculating the LBP histograms at three orthogonal slices (in the case of brain images, we select axial, coronal and sagittal planes) and concatenating them. The resulting feature vector is then only three times larger than the 2D feature vector.

6.3.3 Feature extraction

We subdivide the images into several 3D local patches (patch configuration will be detailed in Section 6.4). The number of patches ranges from a few dozens (in the case of large non-overlapping patches) to some hundred thousands (when we use small highly overlapping patches). We then use the LBP-TOP histogram (LBPH) as the descriptor of each local patch in the image.

Furthermore, because we are interested in localizing the brain regions that are most discriminative in early-stage AD, we follow a multi-classification approach in which we extract the LBPH descriptors at each patch separately. These feature vectors are then fed into individual classifiers to build a classifier ensemble.

Figure 6.2 shows a summary of the feature extraction steps, which ultimately result in one LBPH per 3D image patch.

6.3.4 Classification

To classify each patch, we use a computationally simple classifier called template matching. It consists of determining a "template" for each class from the training samples and selecting the class for which the distance of the patch to each class template is lowest. We use the Chi-square (χ^2) histogram distance measure:

$$\chi^2(p, q) = \frac{1}{2} \sum_{i=1}^{N} \frac{(p_i - q_i)^2}{p_i + q_i} \tag{6.3}$$

with N being the total number of features (histogram bins) and p and q two histograms with N bins.

A popular approach to obtain the final results from a classifier ensemble is to use a fixed combination rule, such as majority voting, since trainable classifier combination rules typically demand large training sets [137], especially when the number of base classifiers (patches) is large [142]. They are also computationally more cumbersome, which is particularly problematic in our application since we use a number of local patches that is in the order of tens of thousands.

Instead of performing simple majority voting, we weigh the prediction of each individual classifier according to its classification accuracy. We propose that the weights be given by a power law with fixed parameter β (with values in the order of $\beta = 10$). In this way, larger weights are assigned to classifiers

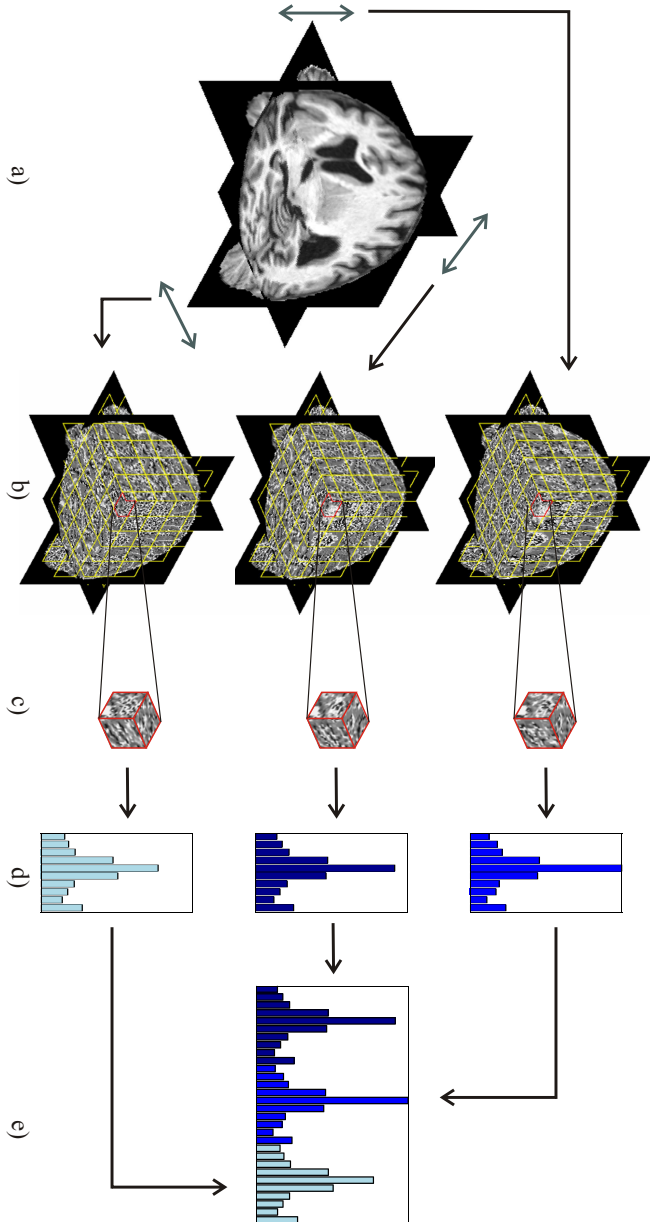Figure 6.2: General overview of the patchwise feature extraction step: a) Subject's 3D image; for all slices in the three orthogonal directions (axial, frontal and sagittal), the LBP labels are computed, originating b) the three LBP label volumes; from each of these volumes, c) local patches are extracted; d) their respective LBP label histograms (LBPH) are computed and e) concatenated to form the final LBPH patch descriptor.

that perform best individually (the best patch is assigned a weight of 1). The commonly used majority vote rule is a particular case of weighted voting, in which the weights are the same for all classifiers.

The decision function is then defined as follows:

$$f(x) = \text{sgn}\left(\frac{\sum_{j=1}^{N} f_j(x)w_j}{\sum_{j=1}^{N} w_j}\right) \tag{6.4}$$

where f_j is the decision function $\{-1, +1\}$ for classifier j and the weight w_j is given by:

$$w_j = \frac{(a_j - 50)^{\beta}}{a_M - 50} \tag{6.5}$$

where a_j and is the accuracy of classifier j and a_M is the accuracy of the best classifier. The accuracies are defined in the range $50\% < a_j \le 100\%$. In the case of 50% accuracy, we set $w_j = 0$.

6.4   Experiments and Results

6.4.1   Data

We use data from the publicly available OASIS (Open Access Series of Imaging Studies) database (`http://www.oasis-brains.org`). The data consist of T1 images that have been previously aligned to a template, corrected for bias field inhomogeneities and brain masked [123].

From the 436 subjects available, we use data from subjects aged 60-80 years old, divided into two classes: 66 normal controls (NC) and 70 subjects with very mild to mild AD. The designation of very mild and mild is based on the Clinical Dementia Rate (CDR). A CDR of 0 corresponds to the normal cognitive state, 0.5 to very mild dementia, 1 to mild dementia, 2 to moderate dementia and 3 to severe dementia [124]. In this database, all subjects with dementia (CDR > 0) had been diagnosed with probable AD.

The demographic information of these subjects is summarized in Table 6.1.

6.4.2   Classification

To evaluate each classifier's generalizability, we perform leave-one-out cross-validation. At each fold, we determine the class templates by averaging the

Table 6.1: Demographic information of the subjects in the two classification classes.

	Normal Controls	Alzheimer's patients
Age	70.8 ± 5.5	73.2 ± 4.8
Gender (F/M)	48/18	41/29
CDR* (0/0.5/1)	66/0/0	0/48/22
MMSE**	29.1 ± 1.1	24.5 ± 4.2

*CDR: Clinical Dementia Rate
**MMSE: Mini-mental State Examination

respective LBPH descriptors. We then determine the distance of the test sample's LBPH to each of the two templates and assign it with the class label that corresponds to the lowest distance.

The sensitivity and specificity are computed at the end of the cross-validation. To avoid overestimated performances due to the classes being slightly imbalanced, we define here classification accuracy as the average between sensitivity and specificity (also known as "balanced accuracy" - bac [143]). This classification is performed separately for all patches in the image.

In the following experiments, we subdivide the images into overlapping patches with voxel sizes of $10\times10\times10$, $20\times20\times20$, $30\times30\times30$ and $40\times40\times40$. These sizes are selected such as to capture local brain regions at increasing scales (Figure 6.3a)). The overlapping ratio (percentage of voxels common to adjacent patches, as depicted in Figure 6.3b)) is 80%. In this way, we are able to obtain a rather high resolution of texture descriptors, which is important for localization purposes. The effect of this ratio on the classification performance is later evaluated.

First, we analyze the performance of the template matching (using χ^2 the histogram distance measure) as an individual classifier. In particular, we determine the histogram of the individual patch accuracies at the four patch sizes. The outputs of the classifiers that perform worse than random (balanced accuracy lower than 50%) are first "negated" (NC becomes AD and vice-versa).

In this experiment, we use $P = 8$ and $r = 1$ for the LBP operator. Later we evaluate the sensitivity of the method to these parameters.

The results are shown in Figure 6.4.

We observe that, as expected, the majority of the patches performs only slightly better than random, with the histograms reaching their peaks, for all
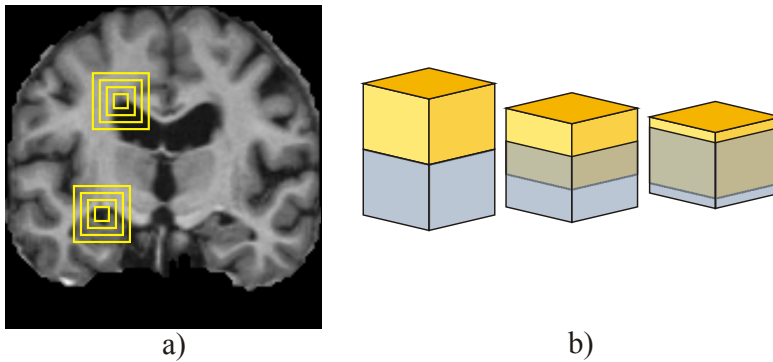
Figure 6.3: Examples of local patches of various sizes (cubes with 10, 20, 30 and 40 side lengths) and overlapping ratios (0%, 50% and 80%). For clarity purposes, two different patches are shown in different colors (blue and yellow) and their overlapping region is shown as a mixture of these.

Figure 6.4: Normalized histogram of the individual patch accuracies for different patch sizes (cube side length, in voxels).

patch sizes, in the range between 50% and 60%. In Table 6.2 we show the results obtained by the best individual patch, for all patch sizes.

Remarkably, the best patch alone achieves high classification accuracies, for all patch sizes, achieving over 87% in the case of the smallest patches. The fact

Table 6.2: Best patch classification results.

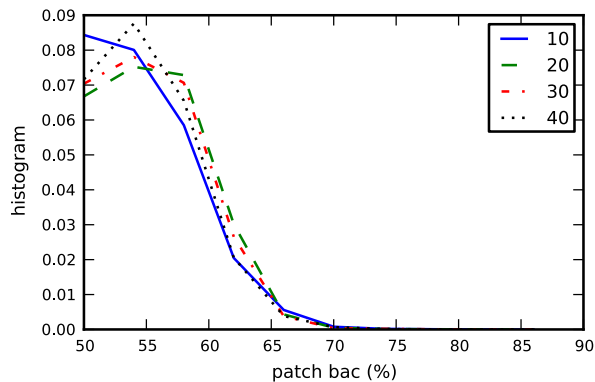| patch size	balanced accuracy (%)	sensitivity (%)	specificity (%)
10×10×10	87.4	90.0	84.8
20×20×20	83.2	78.6	87.9
30×30×30	81.5	84.3	78.8
40×40×40	77.8	81.4	74.2

that higher classification accuracies are obtained with smaller patches might indicate that early changes are contained in small scales of the image.

The following step consists of making use of the various patch classifiers, by combining them to build the final classifier.

Ensemble fusion rule

The simplest fixed combination rule is the majority vote, in which all classification results are combined with the same weight. Table 6.3 shows the results of this combination.

Table 6.3: Simple majority vote classification results.

patch size	balanced accuracy (%)	sensitivity (%)	specificity (%)
10×10×10	75.2	70.0	80.3
20×20×20	69.3	62.9	75.8
30×30×30	67.9	60.0	75.8
40×40×40	66.4	58.6	74.2

When all patches are combined with equal weights, the accuracy drops significantly with respect to that obtained by the best individual patches. This means that the equal inclusion of all patches deteriorates the final classifier performance. This is expected considering that, as shown in Figure 6.4, most patches are non-discriminative. Indeed, we observe that only a small amount of patches contain discriminative information. Therefore, we can intuitively as-

sume that those patches performing well individually should contribute more to the final decision.

We then use a fixed accuracy-weighted rule to combine the classification of the individual patches in the ensemble. This way we avoid an extra training step to find the optimal combination weights, which can be particularly costly when we have hundreds of thousands of patches (necessary in the localization analysis), as is the case of when we use $10\times10\times10$ patches with 80% overlap.

The only parameter of the combination rule is the exponent β in Equation 6.5, which controls the speed at which the weight given to each classifier increases with the classifier's individual accuracy. An example of the obtained weights using three different values for β is shown in Figure 6.5.
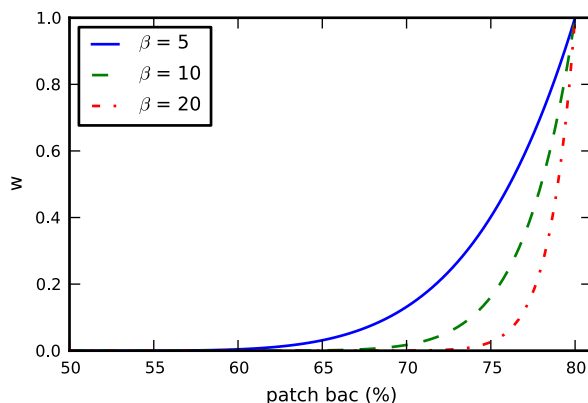
Figure 6.5: Weighting schemes for three different power-law exponents: for lower β, classifiers with lower accuracies are assigned with higher weights than when higher β are used. In this case, the "best" classifier, which receives the weight 1, has an accuracy of 80%.

We evaluate the influence of this parameter in the final classification accuracy and investigate how this approach compares to the single best patch and the commonly used majority vote. As in the previous experiments, we use the LBP$_{8,1}$ operator and patches of four different sizes with 80% overlap subdivision. Figure 6.6 shows the variation of the ensembles' accuracies with respect to β.

The weighted majority vote yields higher classification accuracies than the simple majority voting for all patch sizes, reaching 100% for the smallest patches, at $\beta = 8$. We also observe that the smaller the patches used, the less

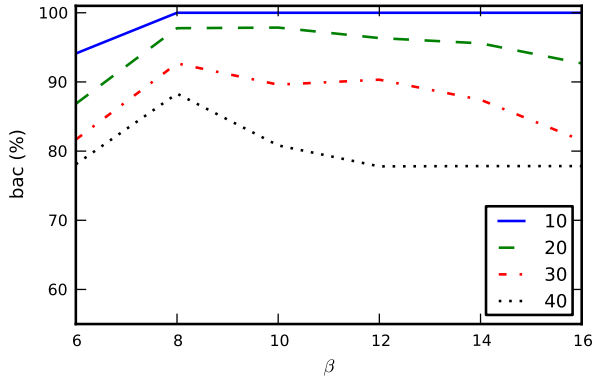Figure 6.6: Evaluation of the influence of the β parameter in the final classification of the weighted voting rule, for four patch sizes (80% overlap).

sensitive the final classifier is to β. In particular, for the $10 \times 10 \times 10$ patches, the accuracy is constant for all $\beta >= 8$. On the other hand, for the largest patches, the performance of the classifier ensemble is only better than that of the "best" classifier for a limited range of β values (lower than 12). In particular, for very large β values, the accuracy of the $40 \times 40 \times 40$ patch ensemble tends to that of the individually best patch (which receives, by definition, a weight of 1).

This can be explained by the fact that we have significantly more patches in the $10 \times 10 \times 10$ (80% overlap) configuration than when larger patches are used. Therefore, the exact weights attributed to each individual patch do not have a significant impact on the final combined performance. In contrast, when only a few discriminative patches are present, a change in the weight given to these patches is likely to influence the final result.

Patch parameters

Subsequently, we evaluate the joint influence of the two patch configuration parameters (size and overlapping ratio) on the performance of the combined classifier. We fix $\beta = 10$.

As shown in Figure 6.7, the accuracy of the ensemble decreases significantly with the patch size. Also, the larger the overlapping ratio (or, in other words, the more information we use to build the ensemble) the better the final classification accuracy. This indicates that, even though some patches might be
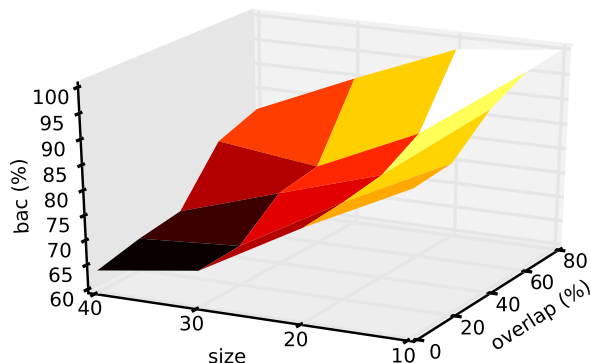
Figure 6.7: Classification results at various patch sizes and overlapping ratios.

highly correlated when the overlapping is large (two neighbor patches contain almost the same information), the overall patch combination still contains complementary discriminative information.

LBP neighborhood parameters

Finally, we evaluate the influence of the number of neighbors, P, and the scale of the LBP operator, r, on the classification accuracy. We use the $10 \times 10 \times 10$ patches with 80% overlap. The results are shown in Figure 6.8.

Smaller scale operators perform better than larger scale ones, with the accuracy decreasing significantly for large radii. Also, for the same radius, a smaller number of neighbors P gives better results, the exception being $r = 1$, with $P = 8$ and $P = 10$ both achieving 100% accuracy. As explained above, the feature vector dimensionality increases with P. Therefore, lower values of P yield smaller feature vectors, helping avoid the small sample problem.

Localization

Besides classification, our other main goal is to localize early-stage AD in the brain. We use the classification accuracies of the individual patches and compute their voxelwise average to obtain the classification maps. This way, these maps indicate which brain regions have a high concentration of highly discriminative local patches. Due to the averaging performed on the patches, a
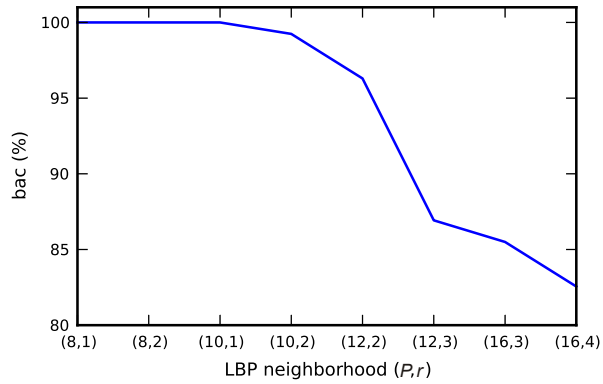
Figure 6.8: Balanced accuracies (bac) at various (P,r) LBP neighborhoods.

spurious highly discriminative patch will not have a high value in these classification maps.
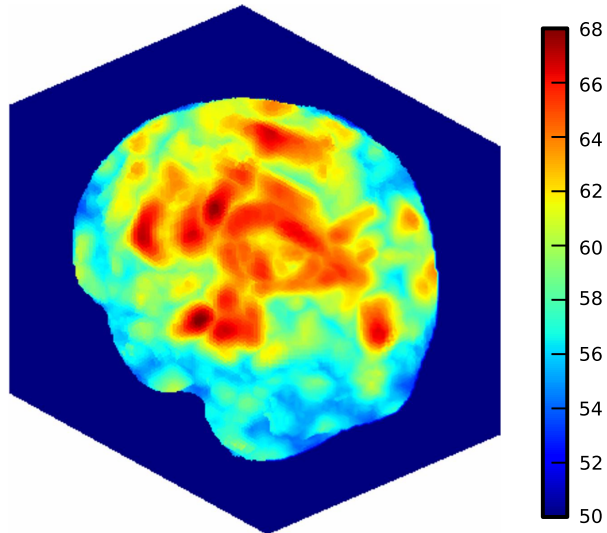

Figure 6.9: Three-dimensional classification map. Each voxel intensity corresponds to the average accuracy of the patches that contain that voxel.

In Figure 6.9 we show the three-dimensional classification map. For visualization purposes, we also show, in Figure 6.10, coronal slices of the classification map at six different locations.

We see that, at an early stage of development, several patches are concentrated near the hippocampi (Figure 6.10c)), with the left one being more discriminative. Several studies have shown asymmetric volume reduction in AD, with the left hippocampus shrinking at a higher rate than the right one [144, 145]. [146] point out that hippocampal asymmetry may be an early sign of the presence of a degenerative process.

The expansion of cerebrospinal fluid (CSF) cavities in the brain is known to occur in early-stage Alzheimer's as a consequence of brain matter loss [147]. Indeed, the ventricles also show as highly discriminative structures in our classification maps. This is in accordance with neuroanatomical findings, which have shown larger ventricular volume in early-stage AD brains compared to that of normal controls [148, 103]. Interhemispheric fissures (Figure 6.10a)-c)), as well as the Sylvian fissure (the right one being highlighted in Figure 6.10c)) are similarly regions of high classification accuracies.

6.4.3   Comparison with other methods

To better place our method among other classification approaches that have been recently proposed to diagnose early-stage AD, we perform comparisons in datasets and cross-validation settings that are as much as possible the same as those used in other works. In particular, we focus on those that have been performed on the OASIS database. An advantage of this database is that the provided images have already been preprocessed (affine registered, bias field corrected and brain masked), eliminating possible differences between steps taken by the various approaches up until the feature extraction step.

In [43], Toews *et al*. train and test (with leave-one-out cross-validation) their method in a cohort of 198 subjects (98 NC and 100 AD). This cohort includes subjects aged 60-96 years, with dementia varying from very mild to moderate. The results are then calculated separately for three sub-cohorts:

1. 66 NC and 22 AD (CDR = 1), age 60-80.

2. 98 NC and 28 AD (CDR = 1), age 60-96.

3. 66 NC and 70 AD (CDR = 0.5 and 1), age 60-80 - the same subdivision we used in the previous experiments.
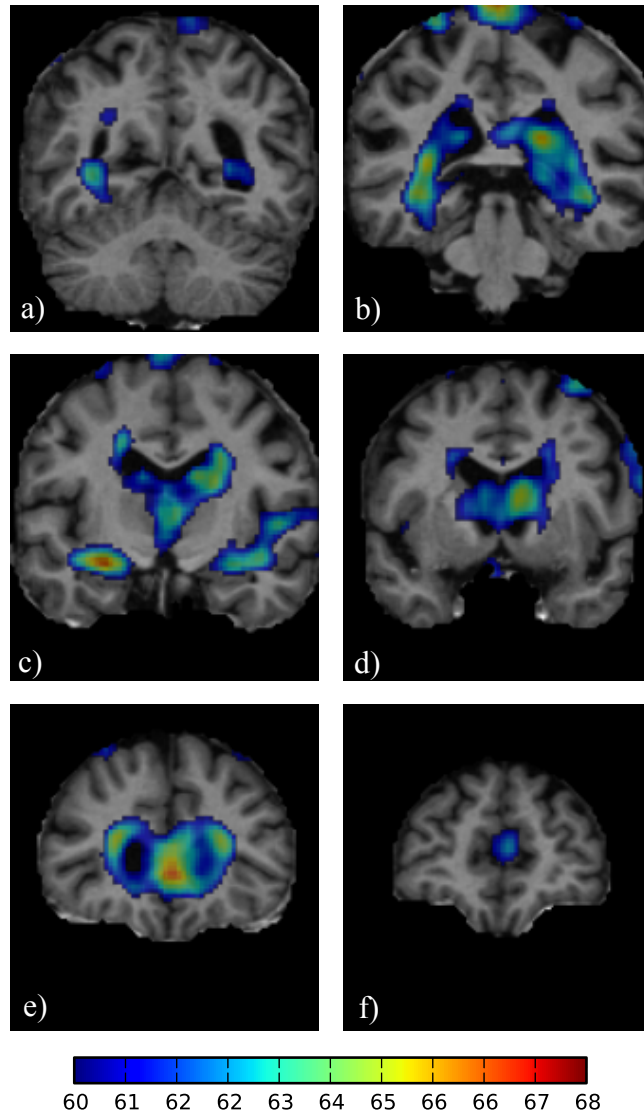
Figure 6.10: Coronal slices of the classification map overlayed on a subject's T1 image, after applying a threshold to the classification map for visualization purposes. a)-f) posterior to anterior.

Similarly, in [131], Daliri proposes the use of SIFT features, followed by classification with Support Vector Machines (SVM). They use the same training-testing configuration as Toews as well as the same subject cohorts. An extension to Toews' method has been proposed by Wang *et al.* [132] to improve feature correspondences.

For a fair comparison with Toews', Daliri's and Wang's methodology, we train and test our method in the exact same subject subdivisions and determine the sensitivity and specificity (in a leave-one-out cross validation setting) for Groups 1, 2 and 3 separately. We use, as previously, the (8,1) LBP descriptor and $10{\times}10{\times}10$ patches with 80% overlap. Results are summarized in Table 6.4.

Table 6.4: Classification results in the literature: SIFT-based approaches.

	1	2	3
[43] (EER*)	80%	70%	71%
[131] (EER*)	86%	78%	75%
[132] (EER*)	80%	79%	71%
ours (bac)	100%	100%	100%

*EER: Equal Error Rate, defined as the value for which the sensitivity and the specificity are the same.

We show that our method outperforms the three SIFT-based approaches presented earlier for the three subdivisions. The differences in performance can be related to several aspects that differ between the two approaches. Firstly, in what concerns the local descriptor used, the SIFT feature vector is determined only at the so-called "interest points", which are located in regions of high intensity gradient. However, the gradients alone can only partially describe the local structure of an image. In particular, two points with the same gradient might have significantly different regional intensity distributions [149]. On the other hand, the LBP descriptor provides a more complete description of the local texture. In addition, we compute these descriptors at the entire brain, not only at specific "interest points", thereby including more information.

Furthermore, [150] proposed a Bag-of-Features (BoF) approach. They extract 2D patches from coronal slices in a training set and build a codebook from those patches. The features used in the classification are then the histograms of the dictionary words present in each image.

They separate the classification tasks into the Groups 1 and 3 described

above. However, they use only one test-training split (70% training - 30% test) for each Group and did not specify which subjects belong to which split. They obtain accuracies of 95% and 85% for the two groups, respectively. For leave-one-out cross validation performed separately in these two Groups, we obtain accuracies of 94% (95% sensitivity, 92.4% specificity) for Group 1 and 100% for Group 3.

Surprisingly, our method performs worse for the subdivision in Group 1 than for Group 3. This would not be expectable, since the AD patients in Group 1 are all at the "mild" stage of the disease (CDR = 1), whereas Group 3 includes brain images from very mild (CDR = 0.5) AD subjects, which are in principle more difficult to classify.

A possible explanation for this is that our method relies on the template matching classifier, which requires representative class templates at each training fold. These might not exist when the number of subjects is low (in this case, only 22 AD samples are present - 21 in each training fold of the leave-one-out cross validation). Another reason might lie in the fact that, for this group, other patch sizes or LBP scales might be more discriminative. In particular, since the AD subjects in this cohort are at a more advanced stage of the disease than those in Group 3, we can expect brain changes to be present at larger scales in the image. Finally, the weighting applied to the patches is also most likely suboptimal. It is possible that, for this subject group, a different accuracy-based weighting scheme should be used.

To compare our results with a volumetric approach on regions of interest, we use data available from the FTP server of the OASIS database website (`ftp://ftp.nrg.wustl.edu/data`). It consists of the volumes of 45 different brain segments, obtained using the widely used open-source software FreeSurfer (`http://surfer.nmr.mgh.harvard.edu`). We use the same group subdivisions as in the comparison with [150], except for 5 subjects in Group 1 and 6 in Group 3, whose data is missing from the volumetry collection.

Using a Linear Discriminant classifier and all 45 volumes as features, the volumetric approach obtains accuracies of 80.5% and 75.0% for Groups 1 and 3, respectively. Finally, we select only the volumes of the left and right hippocampi and amygdalae. The classification accuracies are 70.0% and 70.8% for Groups 1 and 3, respectively (Table 6.5).

This means that, at a very early stage of AD (as in Group 3), the amygdalae and the hippocampi volumes contain the most discriminative information. Adding measurements from other brain structures improves the classification only slightly. In contrast, for Group 1, using the volume of those struc-

Table 6.5: Classification results (percentage) for the volumetric approach (balanced accuracy | sensitivity | specificity).

	1	3
all volumes	80.5 | 70.0 | 90.9	75.0 | 75.7 | 74.2
hippocampi amygdalae	70.0 | 40.0 | 100	70.8 | 64.3 | 77.3
ours	94 | 95.0 | 92.4	100 | 100 | 100

tures alone implies a significant loss of information. A possible cause is that the brain changes are, at this stage of the disease, more spread throughout the brain. Therefore, better performance is obtained when all brain structures are taken into account.

To better visualize the differences between very mild and mild Alzheimer's disease, we analyze the localization results on Group 1 as well. We show the classification map at two coronal slices and compare it with the one obtained for Group 3 (Figure 6.11).

For Group 1, the most significant patches are located, besides the hippocampi and lateral sulci, near the frontal horns of the lateral ventricles. However, for Group 3 these are not as significant, which is consistent with recent volumetric measurements that have been performed on groups of NC, Mild Cognitive Impairment (MCI) and AD. [151] showed a significant difference in the volumes of the frontal horns of the ventricles of NC subjects when compared with those from AD patients. In the NC *vs*. MCI comparison, there was no statistically significant difference.

It is worth noting that although there is not a direct correspondence between the CDR scores in Groups 1 and 3 and the diagnosis of MCI, both MCI subjects and subjects with CDR values of 0.5 and 1 are expected to be largely similar. In particular, a recent study by [15] in a large cohort has shown that, according to the revised criteria for diagnosing MCI [13], 99.8% of the subjects with a CDR score of 0.5 and 93% of those with a CDR of 1 would be classified as MCI.
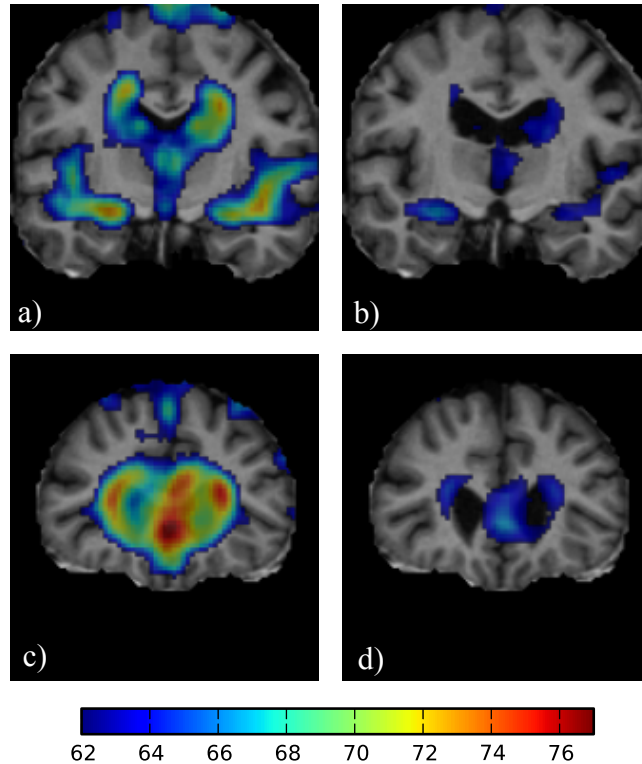
Figure 6.11: Coronal slices of the classification maps for a), c) Group 1 and b), d) Group 3. Note that b) and d) are the same as Figure 6.10c) and e), but shown here in a different color range to facilitate the comparison.

6.5 Conclusion

In this work, we have presented a method for classifying and localizing AD at an early stage of development. We use LBP-TOP descriptors at individual image patches and combine them into a classifier ensemble. We use template matching (with the χ^2 histogram distance) for individual patch classification. This classifier is computationally very simple, therefore allowing for the classification of a large number of 3D patches. In particular, our results indicate that

small highly overlapping patches achieve the highest classification rate when combined in an accuracy-weighted ensemble classifier.

Also, by using a finely sampled image, a rather precise localization is achieved. In particular, we observe that the patches located at the hippocampi and the ventricles are highly discriminative, which is in accordance with recent volumetric and voxel-based studies. This confirms our initial hypothesis that local image patches contain sufficient texture information to be able to discriminate between the two classes. Indeed, 100% classification accuracy is obtained without requiring any a priori assumptions about the location of AD at an early stage and without recurring to complex segmentations and non-linear registrations. One of the purposes of performing non-linear registration to a template is to obtain very accurate inter-subject voxelwise correspondences. One might then expect that by simply affine registering the images the anatomical variability would hinder the classification process. However, the classification maps we obtain show that locations with very high anatomical variability but not necessarily associated with early-stage AD, such as the cortical folds, are indeed not discriminative.

We also demonstrate that the classifier ensemble obtains significantly higher accuracies than the best individual patches, indicating that these are uncorrelated and that different brain regions can provide complementary and discriminative information.

Furthermore, we show that our method outperforms volumetric-based classification. Our classification results are also higher than those obtained by other state-of-the-art local feature approaches. These comparisons have been performed on the exact same subjects and cross-validation configuration.

Future work will include considering other strategies for patch combination and/or selection. In this work, we have proposed a simple weighted majority voting with a fixed parameter. This has the advantage of not needing extra training samples and allowing for an a posteriori combination of patches. However, it is likely that the chosen weights are not optimal for each classification task. This trade-off will have to be more thoroughly evaluated.

Additionally, other patch configurations (and combinations thereof) should be investigated. In particular, it is reasonable to expect that, depending on the local structure of the image, more elongated (parallelepipedic/ellipsoidal) patches can provide better descriptors than the simplistic cube. Similarly, multiscale LBP descriptors should be considered.

Finally, we intend to further evaluate this method's performance in other datasets, preferably from different acquisition centers. More specifically, we are interested in assessing the method's ability to predict conversion to AD.

Conclusion

In this thesis, we have presented methods to help with earlier detection of AD using intensity and texture-based information contained in structural MR images. In this final chapter, we summarize our research results and discuss directions for future work.

7.1 Answers to the research questions

Can we accurately segment white matter hyperintensities from a single MRI modality (FLAIR)?

In Chapter 2, we present a method to automatically segment WMHs using only 3D FLAIR images. We evaluate it in 40 datasets against the ground truth provided by an experienced neuroradiologist. Our results show that the method is suitable for a robust segmentation of WMHs of various loads and that it outperforms other existing unimodal approaches. In particular, we obtain both high similarity and volume correlation coefficients.

   Additionally, we evaluate our proposed method in a publicly available database of Multiple Sclerosis lesions (MICCAI MS Segmentation Challenge), which have a similar appearance as WMH in FLAIR images. The obtained final score (82) indicates that our method performs closely to the human observer, for whom a score of 90 had been given.

Is the proposed method comparable, in terms of performance, to existing multimodal approaches?

Because it does not require additional MRI modalities or atlases, our approach enables the shortening of the acquisition time (thereby increasing patient comfort and potentially reducing motion-related artefacts), avoids the need for

co-registrations and allows for near real-time analysis, making it suitable for
large-scale clinical trials. However, a pertinent question with respect to such
a segmentation method is whether the lack of extra (multimodal) information
compromises its ability to detect lesions. In principle, we might expect that a
trade-off needs to be made between computational complexity and segmenta-
tion accuracy.

To compare our proposed method with existing multimodal approaches,
we apply it on a benchmark database of MS lesions, in which around 30 ap-
proaches have been tested since 2008. Results show that our method performs
similarly to the best performing state-of-the-art multimodal methods, suggest-
ing that either the benefits of using multimodal information are not significant
or they are overshadowed by an increase in the amount of errors that result
from the various registration and normalization steps.

**Is it possible to detect Mild Cognitive Impairment (MCI) using only textural
properties of white matter hyperintensities and what is the performance in
comparison with volumetric/spatial features?**

Our results in a dataset with 14 normal controls and 15 MCI subjects (Chap-
ter 3) show that both intensity histograms and particularly texture descriptors
perform considerably better than random classification when used in a classi-
fication framework to distinguish between these two groups. In contrast, fea-
tures such as the lesion volume and location are not discriminative in the same
dataset. These results indicate that: 1) small-scale (microstructural) differences
exist between normal elderly and MCI subjects at the level of the lesions, cor-
roborating recent findings by diffusion- and perfusion-based studies; 2) these
differences can also be captured by structural MR modalities. For our dataset,
the T2 images are the most discriminative to perform this classification, in com-
parison with the T1 and FLAIR images.

Lesion textures, as captured by T2-based descriptors, are therefore also po-
tential early biomarkers of AD and should be further investigated in large-scale
structural MR imaging studies.

**Do intensity histograms contain enough information to detect early-stage
AD and how do they perform in both traditional and dissimilarity-based
classification frameworks?**

We show in Chapter 4 that image-based features as simple as gray-level his-
tograms achieve about 84% accuracy (73-76% using whole-brain histograms)

when used to discriminate between healthy elderly controls and subjects with very mild to mild AD. Additionally, our results indicate that a dissimilarity-based representation performs better in this classification task than the traditional approach that uses the histograms as feature vectors. Also, our proposed method outperforms that proposed by Klein [120], in which a deformation-based dissimilarity measure is used. These results indicate that, for this dataset, the intensity content of the images is more discriminative than that provided by a large scale measure like the deformation field.

Can texture features help to classify between normal controls and MCI/early-stage AD and how to they perform compared to structural-based features?

In Chapters 5 and 6 we use local texture descriptors to detect MCI and early-stage AD, respectively. In both cases, we show that the texture descriptors not only provide high classification rates but they also perform better than structural-based biomarkers: in the first case, the gray matter concentration map; in the second case, the volumes of the hippocampi and the amygdalae.

Can local patches help in both the classification of MCI/early-stage AD and the localization of the affected brain regions?

Our results presented in Chapters 4, 5 and 6 indicate that patchwise approaches are useful in the classification and localization of MCI/early-stage AD. In particular, the results presented in Chapters 4 and 6 show that patch combination performs better than the individual patches, suggesting that different brain regions contribute positively to the classification. Furthermore, a spatial analysis on the "best" individual patches also reveals which regions are more significant in the discrimination between normal controls and early-stage AD.

Do the detected regions correspond to what is already known about the affected brain structures?

We show in Chapters 5 and 6 that the regions that are significant for classification are located near the hippocampi and the ventricles. Interhemispheric fissures as well as the Sylvian fissure are similarly regions of high classification accuracies. These structures are known, from the medical literature and previous volumetric and morphometric studies, to be affected at an early stage of AD. With these results, we show that it is important to consider the entire

brain in the classification and not only specific priorly segmented regions like the hippocampi. This is in line with the whole-brain voxelwise analyses performed in morphometry studies, with the difference that we do not perform non-linear registrations to obtain inter-subject voxelwise correspondences nor do we rely on prior tissue segmentations.

7.2 Final remarks and recommendations for future work

7.2.1 Segmentation of white matter hyperintensities

Despite the promising results obtained by our proposed lesion segmentation method, it still performs worse than the human observer, as the MICCAI Challenge score seems to indicate. Also, for small lesion loads in the Essen dataset, the correspondence with the manual segmentation by the neuroradiologist is still far from optimal. However, it is worth pointing out that, for this case, an intra- and inter-subject evaluation should be performed in order to have an estimate of the human variability and of how our method compares to it, particularly for low lesion loads.

So far, none of the methods tested on the MICCAI Challenge have performed as well as the human observer. However, a question that remains open, especially when considering the high inter-rater variability, is whether the segmentation performed by the experts on *in vivo* MR images is a real ground truth. In that sense, phantom studies may give a better insight and help towards the development of more accurate and robust segmentation methods.

In any case, several improvements to our method can also be proposed. Firstly, the preprocessing steps play an important role in the subsequent performance of the method. In particular, wrong brain segmentations will likely influence the lesion segmentation, since the skull voxels show also as high intensities in FLAIR images. Likewise, inappropriate bias field corrections will also change the global intensity distributions and therefore influence our method's performance. The effects of varying these two steps need to be quantified and ultimately corrected for, by incorporating them, for example, into the whole segmentation framework.

Finally, in Chapter 2 we use a training set to find the optimal parameters. These are used to extract the binary segmentation from the fuzzy result yielded by the GMM-EM framework. However, a straightforward future implementa-

tion of the method into the clinical setting will enable the online tuning of these parameters according to the clinical judgement by the neuroradiologist.

7.2.2 Texture analysis

In this thesis, we have presented various approaches for texture analysis with the goal of classifying between normal controls and MCI/early-stage AD: 1) texture analysis at the white matter hyperintensities; 2) global and patchwise analysis in affine-registered T1 images using simple intensity histograms; 3) patchwise analysis in non-linearly registered T1 images using second order statistical texture descriptors; 4) patchwise analysis in affine-registered T1 images using Local Binary Pattern descriptors.

We now discuss several aspects of the proposed approaches.

White matter hyperintensities

We show in Chapter 3 that local textures computed only at the white matter hyperintensities are able to capture differences between a group of normal controls and a group of MCI subjects. A drawback of this study that can limit the conclusions drawn from it is the small size of the subject groups. Further research is then needed in a larger dataset, preferably with images acquired at different centers.

However, these preliminary results indicate that white matter hyperintensities should be characterized not only by their volume (which, as stated above, is highly dependent either on the expert or on the automatic segmentation accuracy) or their spatial location, but also by their MR signal properties, particularly textures. Although these microstructural differences have been investigated in perfusion- and diffusion-weighted studies, we show that they are also present in structural MR images.

Furthermore, larger studies are necessary to confirm whether T2 textures are indeed consistently more discriminative than those extracted from T1 and FLAIR images. In particular, since FLAIR is a type of T2-weighted modality (with cerebrospinal fluid signal suppression), it is worth investigating on the differences between FLAIR and T2 lesion properties. Also, a comparison between the lesion textures and the textures computed at normal-appearing white matter may also give some insight into the neurodegeneration processes occurring at an early stage of AD.

Finally, a spatial analysis of which lesions (at which locations) have the

greatest discriminative power between normal controls and MCI is also desirable.

Texture descriptors

In Chapter 5, we use second-order statistical texture descriptors (Haralick features), which are based on pixel pair-wise intensity distributions (the co-occurrence matrix). These descriptors, although often used in medical image analysis, have the disadvantage of requiring a preliminary intensity standardization step. This step is typically performed as a preprocessing step in MR image analysis to make different scans comparable. However, this same intensity matching can have the undesirable consequence of also eliminating disease-related differences between the two groups.

We then advise on the use of local descriptors that are as little sensitive as possible to intensity range differences. In particular, such descriptors are likely to perform better when multicenter datasets are used. In these cases, the MR acquisition parameters are often different, implying that the gray level scales will also be different.

To avoid the intensity standardization step, in Chapters 3 and 6 we propose using local descriptors (LBP histograms) that are insensitive to local intensity differences, as they are based on pairwise intensity comparisons, rather than on their absolute values. The images used in this thesis have been corrected for common MRI artefacts like bias field inhomogeneities and position differences. However, a sensitivity study should be performed to confirm the robustness of the texture descriptors with respect to each individual artefact. These experiments should be performed in images acquired at various MRI centers.

Additionally, even though the literature on texture analysis is extensive, with many proposed descriptors and respective adaptations, our goal in this work was to explore frameworks than can be easily adapted to use other descriptors. Therefore, no thorough comparison between existing texture features has been performed. However, properties such as those exhibited by LBP descriptors are desirable in our application and should therefore be present in further studies of this kind.

It is also worth pointing out that a possible drawback of texture analysis in medical imaging studies is that often no direct medical interpretation can be obtained from the extracted texture features. This is especially true for tissues whose images show no particularly dominant orientations or frequencies, as is the case with brain tissues. Therefore, it happens often that only texture differences (between groups) are analyzed rather than their absolute values.

An interesting direction for further texture analysis is to investigate whether the actual texture values in a certain local patch (or region-of-interest) can be translated into clinical parameters. These can then also be useful in monitoring the progression of the disease, the same way volumetrics has been applied to the study of the evolution of certain brain structures during the course of Alzheimer's disease.

Patchwise approach

Registering individual subject images to a template is an often needed preprocessing step in groupwise MR image analysis. It corrects for differences in patient positioning, thereby enabling local image comparisons. Except for the WMH segmentation method we propose in Chapter 2, which is fully data-driven (meaning that no information from other subjects is required), and the whole-brain DBC framework that uses histograms (these, by definition, lack spatial information), the group comparisons we perform elsewhere require that the subjects' images be aligned. This is particularly true for Chapters 5, 6, and for the local approach presented in Chapter 4, where the localization of affected brain regions is also aimed for.

As explained in Chapter 1, morphometric techniques such as VBM require the images to be non-linearly registered in order for a voxelwise correspondence to be achieved. In Chapter 5 we follow this approach to enable the performance comparison between the textural and the VBM features.

However, and similarly to what has been said about intensity standardization methods, the attempt to align images in a voxel-by-voxel way also comes with the risk of an over-alignment which can result in informative patterns being eliminated from the images. Determining what is acceptable as a final registration result is also not straightforward.

Having that in mind, in Chapters 4 (local approach) and 6 we propose performing simple affine registration, which can correct for positioning and scaling differences but will not interfere with the relative shapes of the brain structures. Additionally, instead of voxelwise features, we use local patchwise features. Our results show that regions of high anatomical variability, but not necessarily associated with AD, such as the cortical folds, perform similarly to random classification and do therefore not interfere with the final localization maps.

Therefore, this patchwise methodology seems promising in both classifying early-stage AD (by combining information from various brain regions) and localizing the regions that contribute most to the classification, without recurring

to segmentations or complex non-linear registrations. However, a study on the robustness of the patchwise approach with respect to minor changes in the affine registration is also desirable. Considering the dimensions of the patches used (more than 10 voxels on each cube side), we can expect than minor registration errors will not significantly hinder the localization and classification processes.

Finally, the use of cubic patches raises the question about the type of information that the intensity-based texture descriptors are able to capture. In particular, two types of groupwise differences can be reflected in the textures: 1) local shape differences which cause differences in the tissue proportions, edge shapes and orientations; 2) local tissue microstructural differences (for example, white matter degeneration). However, the only way to distinguish the second effect from the first one would be to determine the texture descriptors within the voxels of the same tissue. This requires a preliminary tissue segmentation, which we have avoided in this thesis (except in Chapter 3).

In any case, it is worth further investigating whether certain patches are more discriminative because they enclose regions with shape differences (like shrunk hippocampi or enlarged ventricles) or because there are microstructural differences at the tissue level. On the other hand, the fact that no assumptions need to be made about the type of differences between healthy and AD subjects brings the additional advantage of making this approach applicable to other imaging modalities besides structural MRI.

Longitudinal analysis

Besides helping to diagnose AD at an early stage, imaging biomarkers are also desirable to monitor the progression of the disease and the efficacy of the treatments. Both volumetric and morphometric features have been used with this goal. In particular, the rate of hippocampal atrophy has been used to evaluate AD progression [152]. An adapted version of voxel-based morphometry has also been applied to assess gray matter atrophy over time in groups of normal controls and MCI/AD subjects [153].

In this thesis, our focus has been on cross-sectional analyses - images acquired from a population, at one time instant. A natural extension of the proposed methods is their application on longitudinal data - images acquired from a population over time. In particular, instead of using absolute texture descriptor values determined once for each brain image or each local region, a possible extension is to determine (relative) differences between subsequent time instants. In the case of local analysis, these differences can highlight lo-

cations that have changed significantly in texture. Similarly to what is done in cross-sectional studies, these differences (rather than their absolute values determined at specific time instants) can be used in a classification framework to discriminate AD patients from normal controls.

It is worth pointing out that, although useful in the evaluation of the disease progression, in practice, obtaining longitudinal data implies waiting for at least some months to be able to extract meaningfully changing biomarkers, which may be disadvantageous considering the ultimate goal of diagnosing AD at its initial stages.

Bibliography

[1] H. O. Tayeb, H. D. Yang, B. H. Price, and F. I. Tarazi, "Pharmacotherapies for Alzheimer's disease: beyond cholinesterase inhibitors," *Pharmacology & therapeutics*, vol. 134, no. 1, pp. 8–25, 2012.

[2] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of Alzheimer's disease," *Alzheimer's & dementia*, vol. 3, no. 3, pp. 186–191, 2007.

[3] M. Prince, R. Bryce, C. Ferri *et al.*, "World alzheimer report 2011: The benefits of early diagnosis and intervention," *Alzheimer's Disease International*, vol. 15, pp. 5–65, 2011.

[4] S. Akhondzadeh and M. Noroozian, "Alzheimer's disease: pathophysiology and pharmacotherapy." *IDrugs: the investigational drugs journal*, vol. 5, no. 11, pp. 1062–1069, 2002.

[5] J. A. Hardy and G. A. Higgins, "Alzheimer's disease: the amyloid cascade hypothesis." *Science*, 1992.

[6] C. Reitz, "Alzheimer's disease and the amyloid cascade hypothesis: a critical review." *Int J Alzheimers Dis*, vol. 2012, p. 369808, 2012.

[7] H. Braak and E. Braak, "Neuropathological stageing of Alzheimer-related changes," *Acta neuropathologica*, vol. 82, no. 4, pp. 239–259, 1991.

[8] S. Gauthier, B. Reisberg, M. Zaudig, R. C. Petersen, K. Ritchie, K. Broich, S. Belleville, H. Brodaty, D. Bennett, H. Chertkow *et al.*, "Mild cognitive impairment," *The Lancet*, vol. 367, no. 9518, pp. 1262–1270, 2006.

[9] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease." *Nat Rev Neurol*, vol. 6, no. 2, pp. 67–77, Feb 2010.

[10] S. T. DeKosky and K. Marek, "Looking backward to move forward: early detection of neurodegenerative disorders," *Science*, vol. 302, no. 5646, pp. 830–834, 2003.

[11] C. R. Jack Jr, M. S. Albert, D. S. Knopman, G. M. McKhann, R. A. Sperling, M. C. Carrillo, B. Thies, and C. H. Phelps, "Introduction to the recommendations from

the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 3, pp. 257–262, 2011.

[12] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, "Mild cognitive impairment: clinical characterization and outcome." *Arch Neurol*, vol. 56, no. 3, pp. 303–308, Mar 1999.

[13] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux *et al.*, "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging - Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's and Dementia*, vol. 7, no. 3, pp. 263–269, 2011.

[14] M. Ries, C. Carlsson, H. Rowley, M. Sager, C. Gleason, S. Asthana, and S. Johnson, "MRI characterization of brain structure and function in mild cognitive impairment: A review," *J Am Geriatr Soc*, vol. 56, pp. 920–937, 2008.

[15] J. C. Morris, "Revised criteria for mild cognitive impairment may compromise the diagnosis of Alzheimer disease dementia," *Archives of neurology*, vol. 69, pp. 700–708, 2012.

[16] L. Mosconi, M. Brys, L. Glodzik-Sobanska, S. De Santi, H. Rusinek, and M. J. de Leon, "Early detection of Alzheimer's disease using neuroimaging," *Experimental gerontology*, vol. 42, no. 1, pp. 129–138, 2007.

[17] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology*, vol. 34, no. 7, pp. 939–939, 1984.

[18] R. M. Rangayyan, *Biomedical Image Analysis*, M. R. Neuman, Ed. CRC Press - The Biomedical Engineering Series, 2005.

[19] L. Z. Diaz-de Grenu, J. Acosta-Cabronero, J. Pereira, G. Pengas, G. B. Williams, and P. J. Nestor, "MRI detection of tissue pathology beyond atrophy in Alzheimer's disease: Introducing T2-VBM," *Neuroimage*, vol. 56, no. 4, pp. 1946–1953, 2011.

[20] J. V. Hajnal, D. J. Bryant, L. Kasuboski, P. M. Pattany, B. De Coene, P. D. Lewis, J. M. Pennock, A. Oatridge, I. R. Young, and G. M. Bydder, "Use of fluid attenuated inversion recovery (FLAIR) pulse sequences in MRI of the brain," *Journal of computer assisted tomography*, vol. 16, no. 6, pp. 841–844, 1992.

[21] M. Symms, H. Jäger, K. Schmierer, and T. Yousry, "A review of structural magnetic resonance neuroimaging," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, no. 9, pp. 1235–1244, 2004.

[22] W. Chen, X. Song, S. Beyea, R. D'Arcy, Y. Zhang, and K. Rockwood, "Advances in perfusion magnetic resonance imaging in Alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 2, pp. 185–196, 2011.

[23] F. Agosta, M. Pievani, S. Sala, C. Geroldi, S. Galluzzi, G. B. Frisoni, and M. Filippi, "White matter damage in Alzheimer disease and its relationship to gray matter atrophy," *Radiology*, vol. 258, no. 3, pp. 853–863, 2011.

[24] M. G. Naylor, V. A. Cardenas, D. Tosun, N. Schuff, M. Weiner, and A. Schwartzman, "Voxelwise multivariate analysis of multimodality magnetic resonance imaging," *Human brain mapping*, 2013.

[25] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, O. Colliot, and A. D. N. Initiative, "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database." *Neuroimage*, vol. 56, no. 2, pp. 766–781, May 2011.

[26] C. Jack, R. C. Petersen, Y. C. Xu, P. C. O'Brien, G. E. Smith, R. J. Ivnik, B. F. Boeve, S. C. Waring, E. G. Tangalos, and E. Kokmen, "Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment," *Neurology*, vol. 52, no. 7, pp. 1397–1397, 1999.

[27] M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehéricy, H. Benali, L. Garnero, O. Colliot, and A. D. N. I. , "Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI." *Hippocampus*, vol. 19, no. 6, pp. 579–587, Jun 2009.

[28] E. Gerardin, G. Chételat, M. Chupin, R. Cuingnet, B. Desgranges, H.-S. Kim, M. Niethammer, B. Dubois, S. Lehéricy, L. Garnero, F. Eustache, O. Colliot, and A. D. N. I. , "Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging." *Neuroimage*, vol. 47, no. 4, pp. 1476–1486, Oct 2009.

[29] D. P. Devanand, G. Pradhaban, X. Liu, A. Khandji, S. De Santi, S. Segal, H. Rusinek, G. H. Pelton, L. S. Honig, R. Mayeux, Y. Stern, M. H. Tabert, and M. J. de Leon, "Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease." *Neurology*, vol. 68, no. 11, pp. 828–836, Mar 2007.

[30] M. P. Laakso, H. Soininen, K. Partanen, E. L. Helkala, P. Hartikainen, P. Vainio, M. Hallikainen, T. Hänninen, and P. Riekkinen, Sr, "Volumes of hippocampus, amygdala and frontal lobes in the MRI-based diagnosis of early Alzheimer's disease: correlation with memory functions." *J Neural Transm Park Dis Dement Sect*, vol. 9, no. 1, pp. 73–86, 1995.

[31] X. Yang, M. Z. Tan, and A. Qiu, "CSF and brain structural imaging markers of the Alzheimer's pathological cascade," *PloS one*, vol. 7, no. 12, p. e47406, 2012.

[32] O. Querbes, F. Aubry, J. Pariente, J.-A. Lotterie, J.-F. Démonet, V. Duret, M. Puel, I. Berry, J.-C. Fort, P. Celsis *et al.*, "Early diagnosis of Alzheimer's disease using

cortical thickness: impact of cognitive reserve," *Brain*, vol. 132, no. 8, pp. 2036–2047, 2009.

[33] Y. Fan, N. Batmanghelich, C. M. Clark, and C. Davatzikos, "Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline," *Neuroimage*, vol. 39, no. 4, pp. 1731–1743, 2008.

[34] A. Mouiha and S. Duchesne, "Hippocampal atrophy rates in Alzheimer's disease: Automated segmentation variability analysis," *Neuroscience letters*, vol. 495, no. 1, pp. 6–10, 2011.

[35] G. B. Frisoni and C. R. Jack, "Harmonization of magnetic resonance-based manual hippocampal segmentation: a mandatory step for wide clinical use," *Alzheimer's & Dementia*, vol. 7, no. 2, pp. 171–174, 2011.

[36] J. Ashburner and K. J. Friston, *Human Brain Function, 2nd edition*. Academic Press, 2003, ch. 6 - Morphometry.

[37] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease." *Brain*, vol. 131, no. Pt 3, pp. 681–689, Mar 2008.

[38] J. Ashburner and K. J. Friston, "Voxel-based morphometry–the methods." *Neuroimage*, vol. 11, no. 6 Pt 1, pp. 805–821, Jun 2000.

[39] J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsrude, C. Price, and K. Friston, "Identifying global anatomical differences: deformation-based morphometry." *Hum Brain Mapp*, vol. 6, no. 5-6, pp. 348–357, 1998.

[40] S. J. Teipel, C. Born, M. Ewers, A. L. W. Bokde, M. F. Reiser, H.-J. Möller, and H. Hampel, "Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment." *Neuroimage*, vol. 38, no. 1, pp. 13–24, Oct 2007.

[41] J. Koikkalainen, J. Lötjönen, L. Thurfjell, D. Rueckert, G. Waldemar, and H. Soininen, "Multi-template tensor-based morphometry: application to analysis of Alzheimer's disease," *NeuroImage*, vol. 56, no. 3, pp. 1134–1144, 2011.

[42] F. L. Bookstein, "Voxel-based morphometry should not be used with imperfectly registered images." *Neuroimage*, vol. 14, no. 6, pp. 1454–1462, 2001.

[43] M. Toews, W. Wells, 3rd, D. L. Collins, and T. Arbel, "Feature-based morphometry: discovering group-related anatomical patterns." *Neuroimage*, vol. 49, no. 3, pp. 2318–2327, Feb 2010.

[44] C. Davatzikos and S. M. Resnick, "Degenerative age changes in white matter connectivity visualized in vivo using magnetic resonance imaging." *Cereb Cortex*, vol. 12, no. 7, pp. 767–771, Jul 2002.

[45] R. V. Parsey and K. Krishnan, "Quantitative analysis of t2 signal intensities in alzheimer's disease," *Psychiatry Research: Neuroimaging*, vol. 82, no. 3, pp. 181–185, 1998.

[46] A. Kassner and R. E. Thornhill, "Texture analysis: a review of neurologic mr imaging applications." *AJNR Am J Neuroradiol*, vol. 31, no. 5, pp. 809–816, May 2010.

[47] V. A. Kovalev, F. Kruggel, H.-J. Gertz, and D. Y. von Cramon, "Three-dimensional texture analysis of MRI brain datasets," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 5, pp. 424–433, 2001.

[48] P. A. Freeborough and N. C. Fox, "MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease," *IEEE Trans Med Imaging*, vol. 17, no. 3, pp. 475–478, 1998.

[49] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, "Texture analysis of medical images." *Clin Radiol*, vol. 59, no. 12, pp. 1061–1069, Dec 2004.

[50] Y. Zhang, "MRI texture analysis in multiple sclerosis," *Journal of Biomedical Imaging*, vol. 2012, p. 2, 2012.

[51] J. Rajeesh, "Discrimination of Alzheimer's disease using hippocampus texture features from MRI," *Asian Biomedicine (Research Reviews and News)*, vol. 6, no. 1, 2012.

[52] J. Zhang, C. Yu, G. Jiang, W. Liu, and L. Tong, "3D texture analysis on MRI images of Alzheimer's disease." *Brain Imaging Behav*, vol. 6, no. 1, pp. 61–69, Mar 2012.

[53] M. S. de Oliveira, M. L. F. Balthazar, A. D'Abreu, C. L. Yasuda, B. P. Damasceno, F. Cendes, and G. Castellano, "MR imaging texture analysis of the corpus callosum and thalamus in amnestic Mild Cognitive Impairment and mild Alzheimer disease." *AJNR Am J Neuroradiol*, vol. 32, no. 1, pp. 60–66, Jan 2011.

[54] L. Pantoni and J. H. Garcia, "The significance of cerebral white matter abnormalities 100 years after Binswanger's report. A review." *Stroke*, vol. 26, no. 7, pp. 1293–1301, Jul 1995.

[55] F. de Leeuw, J. C. de Groot, E. Achten, M. Oudkerk, L. Ramos, R. Heijboer, A. Hofman, J. Jolles, G. van Gijn, and M. M Breteler, "Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study. the Rotterdam Scan Study," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 1, pp. 9–14, 2001.

[56] L. J. Launer, K. Berger, M. M. B. Breteler, C. Dufouil, R. Fuhrer, S. Giampaoli, L.-G. Nilsson, A. Pajak, M. de Ridder, E. J. van Dijk, S. Sans, R. Schmidt, and A. Hofman, "Regional variability in the prevalence of cerebral white matter lesions: an MRI study in 9 European countries (CASCADE)." *Neuroepidemiology*, vol. 26, no. 1, pp. 23–29, 2006.

[57] A. Ylikoski, T. Erkinjuntti, R. Raininko, S. Sarna, R. Sulkava, and R. Tilvis, "White matter hyperintensities on MRI in the neurologically nondiseased elderly. Analysis of cohorts of consecutive subjects aged 55 to 85 years living at home." *Stroke*, vol. 26, no. 7, pp. 1171–1177, Jul 1995.

[58] R. O. Hopkins, C. J. Beck, D. L. Burnett, L. K. Weaver, J. Victoroff, and E. D. Bigler, "Prevalence of white matter hyperintensities in a young healthy population." *Journal of Neuroimaging*, vol. 3, pp. 243–251, 2006.

[59] S. Debette and H. S. Markus, "The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis." *BMJ*, vol. 341, p. c3666, 2010.

[60] E. Olsson, N. Klasson, J. Berge, C. Eckerström, Å. Edman, H. Malmgren, and A. Wallin, "White matter lesion assessment in patients with cognitive impairment and healthy controls: Reliability comparisons between visual rating, a manual, and an automatic volumetrical MRI method - the Gothenburg MCI study," *Journal of aging research*, vol. 2013, 2013.

[61] L. C. Silbert, H. H. Dodge, L. G. Perkins, L. Sherbakov, D. Lahna, D. Erten-Lyons, R. Woltjer, L. Shinto, and J. A. Kaye, "Trajectory of white matter hyperintensity burden preceding mild cognitive impairment," *Neurology*, vol. 79, no. 8, pp. 741–747, 2012.

[62] M. Mortamais, C. Reynes, A. M. Brickman, F. A. Provenzano, J. Muraskin, F. Portet, C. Berr, J. Touchon, A. Bonafé, E. le Bars *et al.*, "Spatial distribution of cerebral white matter lesions predicts progression to mild cognitive impairment and dementia," *PloS one*, vol. 8, no. 2, p. e56972, 2013.

[63] C. DeCarli, D. Mungas, D. Harvey, B. Reed, M. Weiner, H. Chui, and W. Jagust, "Memory impairment, but not cerebrovascular disease, predicts progression of MCI to dementia." *Neurology*, vol. 63, pp. 220–227, 2004.

[64] R. Schmidt, S. Ropele, C. Enzinger, K. Petrovic, S. Smith, H. Schmidt, P. M. Matthews, and F. Fazekas, "White matter lesion progression, brain atrophy, and cognitive decline: the Austrian stroke prevention study," *Annals of neurology*, vol. 58, no. 4, pp. 610–616, 2005.

[65] E. D. Ross, S. L. Hansel, D. M. Orbelo, and M. Monnot, "Relationship of leukoaraiosis to cognitive decline and cognitive aging," *Cognitive and behavioral neurology*, vol. 18, no. 2, p. 89, 2005.

[66] I. Makedonov, S. Black, and B. MacIntosh, "Cerebral small vessel disease in aging and Alzheimer's disease: a comparative study using MRI and SPECT," *European Journal of Neurology*, vol. 20, no. 2, pp. 243–250, 2013.

[67] K. Oppedal, K. Engan, D. Aarsland, M. Beyer, O. Tysnes, and T. Eftestol, "Using local binary pattern to classify dementia in MRI," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. IEEE, 2012, pp. 594–597.

[68] P. Scheltens, T. Erkinjunti, D. Leys, L. O. Wahlund, D. Inzitari, T. del Ser, F. Pasquier, F. Barkhof, R. Mäntylä, J. Bowler, A. Wallin, J. Ghika, F. Fazekas, and L. Pantoni, "White matter changes on CT and MRI: an overview of visual rating scales. European task force on age-related white matter changes." *Eur Neurol*, vol. 39, no. 2, pp. 80–89, 1998.

[69] R. Mäntylä, T. Erkinjuntti, O. Salonen, H. Aronen, T. Peltonen, T. Pohjasvaara, and C. Standertskjöld-Nordensta, "Variable agreement between visual rating scales for white matter hyperintensities on MRI: comparison of 13 rating scales in a poststroke cohort," *Stroke*, vol. 28, pp. 1614–1623, 1997.

[70] E. C. W. van Straaten, F. Fazekas, E. Rostrup, P. Scheltens, R. Schmidt, L. Pantoni, D. Inzitari, G. Waldemar, T. Erkinjuntti, R. Mäntylä, L.-O. Wahlund, F. Barkhof, and L. A. D. I. S. Group, "Impact of white matter hyperintensities scoring method on correlations with clinical data: the LADIS study." *Stroke*, vol. 37, no. 3, pp. 836–840, Mar 2006.

[71] D. M. J. van den Heuvel, V. H. ten Dam, A. J. M. de Craen, F. Admiraal-Behloul, A. C. G. M. van Es, W. M. Palm, A. Spilt, E. L. E. M. Bollen, G. J. Blauw, L. Launer, R. G. J. Westendorp, M. A. van Buchem, and P. R. O. S. P. E. R. S. Group, "Measuring longitudinal white matter changes: comparison of a visual rating scale with a volumetric measurement." *AJNR Am J Neuroradiol*, vol. 27, no. 4, pp. 875–878, Apr 2006.

[72] J. Grimaud, M. Lai, J. Thorpe, P. Adeleine, L. Wang, G. J. Barker, D. L. Plummer, P. S. Tofts, W. I. McDonald, and D. H. Miller, "Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques." *Magn Reson Imaging*, vol. 14, no. 5, pp. 495–505, 1996.

[73] P. Anbeek, K. L. Vincken, M. J. P. van Osch, R. H. C. Bisschops, and J. van der Grond, "Automatic segmentation of different-sized white matter lesions by voxel probability estimation." *Med Image Anal*, vol. 8, no. 3, pp. 205–215, Sep 2004.

[74] F. Admiraal-Behloul, D. M. J. van den Heuvel, H. Olofsen, M. J. P. van Osch, J. van der Grond, M. A. van Buchem, and J. H. C. Reiber, "Fully automatic segmentation of white matter hyperintensities in MR images of the elderly." *Neuroimage*, vol. 28, no. 3, pp. 607–617, Nov 2005.

[75] R. de Boer, H. A. Vrooman, F. van der Lijn, M. W. Vernooij, M. A. Ikram, A. van der Lugt, M. M. B. Breteler, and W. J. Niessen, "White matter lesion extension to automatic brain tissue segmentation on MRI." *Neuroimage*, vol. 45, no. 4, pp. 1151–1161, May 2009.

[76] C. R. Jack, P. C. O'Brien, D. W. Rettman, M. M. Shiung, Y. Xu, R. Muthupillai, A. Manduca, R. Avula, and B. J. Erickson, "FLAIR histogram segmentation for measurement of leukoaraiosis volume." *J Magn Reson Imaging*, vol. 14, no. 6, pp. 668–676, Dec 2001.

[77] A. Khademi, A. Venetsanopoulos, and A. R. Moody, "Robust white matter lesion segmentation in FLAIR MRI." *IEEE Trans Biomed Eng*, vol. 59, no. 3, pp. 860–871, Mar 2012.

[78] N. Shiee, P.-L. Bazin, A. Ozturk, D. S. Reich, P. A. Calabresi, and D. L. Pham, "A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions." *Neuroimage*, vol. 49, no. 2, pp. 1524–1535, Jan 2010.

[79] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi, and N. Ayache, "Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images." *Neuroimage*, vol. 57, no. 2, pp. 378–390, Jul 2011.

[80] F. Barkhof and R. Smithuis. (2007) The Radiology Assistant - Multiple Sclerosis.

[81] E. H. Herskovits, R. Itoh, and E. R. Melhem, "Accuracy for detection of simulated lesions: comparison of fluid-attenuated inversion-recovery, proton density–weighted, and T2-weighted synthetic brain MR imaging." *AJR Am J Roentgenol*, vol. 176, no. 5, pp. 1313–1318, May 2001.

[82] K. Ong, D. Ramachandrama, R. Mandavaa, and I. Shuaibb, "Automatic white matter lesion segmentation using an adaptive outlier detection method," *Magnetic Resonance Imaging*, vol. 30, pp. 807–823, 2012.

[83] K. Blekas, A. Likas, N. P. Galatsanos, and I. E. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Transactions on Neural Networks*, vol. 16, no. 2, pp. 494–498, 2005.

[84] H. Tang, J.-L. Dillenseger, X. D. Bao, and L. M. Luo, "A vectorial image soft segmentation method based on neighborhood weighted Gaussian mixture model." *Comput Med Imaging Graph*, vol. 33, no. 8, pp. 644–650, Dec 2009.

[85] M. Neema, Z. D. Guss, J. M. Stankiewicz, A. Arora, B. C. Healy, and R. Bakshi, "Normal findings on brain fluid-attenuated inversion recovery MR images at 3T." *AJNR Am J Neuroradiol*, vol. 30, no. 5, pp. 911–916, May 2009.

[86] S. Ruan, C. Jaggi, J. Xue, J. Fadili, and D. Bloyet, "Brain tissue classification of magnetic resonance images using partial volume modeling." *IEEE Trans Med Imaging*, vol. 19, no. 12, pp. 1179–1187, Dec 2000.

[87] L. Lemieux, A. Hammers, T. Mackinnon, and R. S. N. Liu, "Automatic segmentation of the brain and intracranial cerebrospinal fluid in T1-weighted volume MRI scans of the head, and its application to serial cerebral and intracranial volumetry." *Magn Reson Med*, vol. 49, no. 5, pp. 872–884, May 2003.

[88] H. Caillol, W. Pieczynski, and A. Hillion, "Estimation of fuzzy Gaussian mixture and unsupervised statistical image segmentation," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 425–440, 1997.

[89] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, No. 1, pp. 1–38, 1977.

[90] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for Gaussian mixtures," *Technical Report, Center for Biological and Computational Learning and the Artificial Intelligence, Laboratory of the Massachusetts Institute of Technology*, 1994.

[91] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Pearson International Edition, 2008.

[92] S. M. Smith, "Fast robust automated brain extraction." *Human Brain Mapping*,, vol. 17, no. 3, pp. 143–155, November 2002.

[93] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm." *IEEE Trans Med Imaging*, vol. 20, no. 1, pp. 45–57, Jan 2001.

[94] J. J. Bartko, "Measurement and reliability: statistical thinking considerations." *Schizophr Bull*, vol. 17, no. 3, pp. 483–489, 1991.

[95] T. B. Dyrby, E. Rostrup, W. F. Baaré, E. C. van Straaten, F. Barkhof, H. Vrenken, S. Ropele, R. Schmidt, T. Erkinjuntti, L. Wahlund, L. Pantonih, D. Inzitari, O. B. Paulson, L. K. Hansen, and G. Waldemar, "Segmentation of age-related white matter changes in a clinical multi-center study," *Neuroimage*, vol. 41, Issue 2, pp. 335–345, 2008.

[96] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 34, pp. 1907–1916, 2004.

[97] K. van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, "Automated segmentation of Multiple Sclerosis lesions by model outlier detection," *IEEE Transactions on Medical Imaging*, vol. 20, pp. 677–688, 2001.

[98] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 381–396, 2002.

[99] M. Styner, J. Lee, B. Chin, M. Chin, O. Commowick, H. Tran, S. Markovic-Plese, V. Jewells, and S. Warfield, "3D segmentation in the clinic: A Grand Challenge II: MS lesion segmentation," 11 2008.

[100] X. Tomas-Fernandez and S. K. Warfield, "A new classifier feature space for an improved Multiple Sclerosis lesion segmentation," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, 2011.

[101] R. C. Petersen, J. C. Stevens, M. Ganguli, E. G. Tangalos, J. L. Cummings, and S. T. DeKosky, "Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology." *Neurology*, vol. 56, no. 9, pp. 1133–1142, May 2001.

[102] J. C. Morris, M. Storandt, J. P. Miller, D. W. McKeel, J. L. Price, E. H. Rubin, and L. Berg, "Mild cognitive impairment represents early-stage Alzheimer disease," *Archives of neurology*, vol. 58, no. 3, p. 397, 2001.

[103] S. M. Nestor, R. Rupsingh, M. Borrie, M. Smith, V. Accomazzi, J. L. Wells, J. Fogarty, R. Bartha *et al.*, "Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database," *Brain*, vol. 131, no. 9, pp. 2443–2454, 2008.

[104] D. Medina, L. DeToledo-Morrell, F. Urresta, J. D. E. Gabrieli, M. Moseley, D. Fleischman, D. A. Bennett, S. Leurgans, D. A. Turner, and G. T. Stebbins, "White matter changes in mild cognitive impairment and ad: A diffusion tensor imaging study." *Neurobiol Aging*, vol. 27, no. 5, pp. 663–672, May 2006.

[105] J. Liu, C. Yin, S. Xia, L. Jia, Y. Guo, Z. Zhao, X. Li, Y. Han, and J. Jia, "White matter changes in patients with amnestic mild cognitive impairment detected by Diffusion Tensor Imaging," *PloS one*, vol. 8, no. 3, p. e59440, 2013.

[106] J. Latoo and F. Jan, "Dementia with Lewy bodies: clinical review," *British Journal of Medical Practitioners*, vol. 1, no. 1, pp. 10–14, 2008.

[107] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, "Local Binary Patterns for still images," *Computer Vision Using Local Binary Patterns*, pp. 13–47, 2011.

[108] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[109] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns," *Computer Vision-ECCV 2000*, pp. 404–420, 2000.

[110] T. Ojala, M. Pietikäinen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[111] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[112] V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273–295, 1995.

[113] A. C. Evans, D. L. Collins, S. Mills, E. Brown, R. Kelly, and T. M. Peters, "3D statistical neuroanatomical models from 305 mri volumes," in *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.* IEEE, 1993, pp. 1813–1817.

[114] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4itk: Improved n3 bias correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[115] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and D. E., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[116] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization." *IEEE Trans Med Imaging*, vol. 19, no. 2, pp. 143–150, Feb 2000.

[117] A. Wimo and M. Prince, "World Alzheimer Report 2010 - The Global Economic Impact of Dementia," Alzheimer's Disease International, Tech. Rep., 2010.

[118] A. Alzheimer, "About a peculiar disease of the cerebral cortex. (translated by l. jarvik and h. greenson)," *Alzheimer Dis Assoc Disord*, vol. 1, no. 1, pp. 3–8, 1907.

[119] H. I. L. Jacobs, M. P. J. Van Boxtel, J. Jolles, F. R. J. Verhey, and H. B. M. Uylings, "Parietal cortex matters in Alzheimer's disease: an overview of structural, functional and metabolic findings." *Neurosci Biobehav Rev*, vol. 36, no. 1, pp. 297–309, Jan 2012.

[120] S. Klein, M. Loog, F. van der Lijn, T. den Heijer, A. Hammers, M. de Bruijne, A. van der Lugt, R. P. Duin, M. M. Breteler, and W. J. Niessen, "Early diagnosis of dementia based on intersubject whole-brain dissimilarities," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*.   IEEE, 2010, pp. 249–252.

[121] E. Pękalska and R. P. Duin, *The dissimilarity representation for pattern recognition: foundations and applications*.   World Scientific, 2005, no. 64.

[122] A. Ulaş, R. P. Duin, U. Castellani, M. Loog, P. Mirtuono, M. Bicego, V. Murino, M. Bellani, S. Cerruti, M. Tansella *et al.*, "Dissimilarity-based detection of schizophrenia," *International Journal of Imaging Systems and Technology*, vol. 21, no. 2, pp. 179–192, 2011.

[123] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults." *J Cogn Neurosci*, vol. 19, no. 9, pp. 1498–1507, Sep 2007.

[124] J. C. Morris, "The Clinical Dementia Rating (CDR): current version and scoring rules." *Neurology*, 1993.

[125] S. Klein, M. Staring, K. Murphy, M. A. Viergever, J. P. Pluim *et al.*, "Elastix: a toolbox for intensity-based medical image registration," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[126] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, pp. 610–621, 1973.

[127] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. G. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods." *J Magn Reson Imaging*, vol. 27, no. 4, pp. 685–691, Apr 2008.

[128] J. Ashburner, "A fast diffeomorphic image registration algorithm." *Neuroimage*, vol. 38, no. 1, pp. 95–113, Oct 2007.

[129] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick, "Detection of prodromal alzheimer's disease via pattern classification of magnetic resonance imaging," *Neurobiology of aging*, vol. 29, no. 4, pp. 514–523, 2008.

[130] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2. Ieee, 1999, pp. 1150–1157.

[131] M. R. Daliri, "Automated diagnosis of Alzheimer disease using the scale-invariant feature transforms in magnetic resonance images." *J Med Syst*, vol. 36, no. 2, pp. 995–1000, Apr 2012.

[132] Y. Wang, J. A. Catindig, S. Hilal, H. W. Soon, E. Ting, T. Y. Wong, N. Venketasubramanian, C. Chen, and A. Qiu, "Multi-stage segmentation of white matter hyperintensity, cortical and lacunar infarcts." *Neuroimage*, vol. 60, no. 4, pp. 2379–2388, May 2012.

[133] S. Marcel, Y. Rodriguez, and G. Heusch, "On the recent use of local binary patterns for face authentication," *International Journal on Image and Video Processing; Special Issue on Facial Image Processing*, pp. 06–34, 2007.

[134] D. Unay, A. Ekin, M. Cetin, R. Jasinschi, and A. Ercil, "Robustness of local binary patterns in brain MR image analysis." *Conf Proc IEEE Eng Med Biol Soc*, vol. 2007, pp. 2098–2101, 2007.

[135] D. Unay and A. Ekin, "Intensity versus texture for medical image search and retrival," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*. IEEE, 2008, pp. 241–244.

[136] C.-W. Chang, C.-C. Ho, and J.-H. Chen, "ADHD classification by a texture analysis of anatomical brain MRI data." *Front Syst Neurosci*, vol. 6, p. 66, 2012.

[137] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2617–2628, 2007.

[138] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[139] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 3, pp. 226–239, 1998.

[140] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Machine Learning: Proceedings of the Thirteenth International Conference*. MORGAN KAUFMANN PUBLISHERS, INC., 1996, pp. 148–156.

[141] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu, "Boosting local binary pattern (LBP)-based face recognition," in *Advances in biometric person authentication*. Springer, 2005, pp. 179–186.

[142] R. P. Duin, "The combining classifier: to train or not to train?" in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 2. IEEE, 2002, pp. 765–770.

[143] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 3121–3124.

[144] H. Wolf, M. Grunwald, F. Kruggel, S. Riedel-Heller, S. Angerhöfer, A. Hojjatoleslami, A. Hensel, T. Arendt, and H.-J. Gertz, "Hippocampal volume discriminates between normal cognition; questionable and mild dementia in the elderly," *Neurobiology of aging*, vol. 22, no. 2, pp. 177–186, 2001.

[145] A. Vijayakumar and A. Vijayakumar, "Comparison of hippocampal volume in dementia subtypes," *ISRN Radiology*, vol. 2013, 2012.

[146] N. Fox, E. Warrington, P. Freeborough, P. Hartikainen, A. Kennedy, J. Stevens, and M. N. Rossor, "Presymptomatic hippocampal atrophy in Alzheimer's disease: A longitudinal MRI study," *Brain*, vol. 119, no. 6, pp. 2001–2007, 1996.

[147] R. I. Scahill, J. M. Schott, J. M. Stevens, M. N. Rossor, and N. C. Fox, "Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI," *Proceedings of the National Academy of Sciences*, vol. 99, no. 7, pp. 4703–4707, 2002.

[148] C. DeCarli, J. Haxby, J. Gillette, D. Teichberg, S. Rapoport, and M. Schapiro, "Longitudinal changes in lateral ventricular volume in datients with dementia of the Alzheimer type," *Neurology*, vol. 42, no. 10, pp. 2029–2029, 1992.

[149] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, "Discriminative local binary patterns for human detection in personal album," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[150] A. Rueda, J. Arevalo, A. Cruz, E. Romero, and F. González, "Bag of features for automatic classification of Alzheimer's Disease in Magnetic Resonance images," *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp. 559–566, 2012.

[151] L. G. Apostolova, A. E. Green, S. Babakchanian, K. S. Hwang, Y.-Y. Chou, A. W. Toga, and P. M. Thompson, "Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment (MCI), and Alzheimer disease," *Alzheimer Disease & Associated Disorders*, vol. 26, no. 1, p. 17, 2012.

[152] J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, C. Avedissian, S. K. Madsen, N. Parikshak, X. Hua, A. W. Toga, C. R. Jack Jr *et al.*, "Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls," *Neuroimage*, vol. 43, no. 1, pp. 59–68, 2008.

[153] G. Chetelat, B. Landeau, F. Eustache, F. Mezenge, F. Viader, V. de La Sayette, B. Desgranges, and J.-C. Baron, "Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study," *Neuroimage*, vol. 27, no. 4, pp. 934–946, 2005.

About the author

Rita Simões was born in Viseu, Portugal, in 1987. She received her BSc and MSc in Biomedical Engineering from Instituto Superior Técnico (Lisbon, Portugal) in 2007 and 2009, respectively. The final year of her master's was spent at the University of Twente (Enschede, The Netherlands) as an exchange student. During this period, she worked on her MSc thesis, at the Biomedical Signals and Systems group. In September 2009 she started her PhD at the same University, in the Signals and Systems group. In December 2013, she will start working as a Post-Doc researcher in medical imaging at the INRIA-Morpheme team, in Sophia Antipolis, France.

List of Publications

- Lopes Simoes, A.R. and van Cappellen van Walsum, A.-M. and Slump, C.H., "Classification and localization of early Alzheimer's disease in Magnetic Resonance images using Local Binary Patterns in a patch-based classifier ensemble." (submitted) 2013

- Luchtenberg, A. and Lopes Simoes, A.R. and van Cappellen van Walsum, A.-M. and Slump, C.H., "Early detection of Alzheimer's disease using MR image histograms in a dissimilarity-based classification framework." Accepted for publication at "SPIE - Medical Imaging: Computer-Aided Diagnosis 2014, San Diego, California, USA"

- Lopes Simoes, A.R. and Mönninghoff, C. and Dlugaj, M. and Weimar, C. and Wanke, I. and van Cappellen van Walsum, A.-M. and Slump, C.H., "Automatic segmentation of cerebral white matter hyperintensities using only 3D FLAIR images." *Magnetic Resonance Imaging*, vol. 31, no. 7, pp. 1182-1189, 2013

- Lopes Simoes, A.R. and Slump, C.H. and van Cappellen van Walsum, A.-M., "Using local texture maps of brain MR images to detect Mild Cognitive Impairment." In: *21st International Conference on Pattern Recognition (ICPR 2012)*, 11-15 Nov 2012, Tsukuba, Japan. pp. 153-156.

- Lopes Simoes, A.R. and Slump, C.H. and Mönninghoff, C. and Wanke, I. and Dlugaj, M. and Weimar, C., "Automatic histogram-based segmentation of white matter hyperintensities using 3D FLAIR images." In: *SPIE Medical Imaging 2012: Computer-Aided Diagnosis*, 4-9 Feb 2012, San Diego, California, USA.

- Lopes Simoes, A.R. and Slump, C.H., "Change detection and classification in brain MR images using Change Vector Analysis." In: *33rd Annual International Conference of the IEEE EMBS 2011*, 30 Aug - 03 Sep 2011, Boston, Massachusetts USA. pp. 7803-7807.

- Maan, B. and Lopes Simoes, A.R. and Meijer, F.J.A. and Renema, W.K.J. and Slump, C.H., "Comparison of dynamic susceptibility contrast-MRI perfusion quantification methods in the presence of delay and dispersion." In: *SPIE Medical Imaging 2011: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 14-17 Feb 2011, Lake Buena Vista, Florida, USA.

Acknowledgments

Years of observational studies have led me to conclude that the Acknowledgements is the most read part of a PhD thesis. However, it is the section which we spend the least amount of time writing. That being said, I hope this will not disappoint you - if it does, please have my previous statement in mind. I would first like to take this opportunity to let you all know that this thesis presents methods to detect Alzheimer's disease at an early stage of development by analyzing the textures of Magnetic Resonance images. If you thought you could get away with avoiding some thesis content by jumping straight into the Acknowledgements section, it is too late now.

The time has then come to acknowledge the help of those people without whom my PhD trajectory would not have reached a happy ending.

I will start by thanking my promoter and supervisor, Kees Slump, for the freedom he gave me and for trusting that I would eventually find out what to do with it; also for the patience in revising the papers and in providing constructive criticism. Secondly, I thank my assistant promoter, Anne-Marie van Cappellen van Walsum, for her precious medical perspective and for always helping me keep a solid bridge between the medical and the technical worlds.

I would also like to acknowledge the cooperation with the doctors at the Essen University Hospital: Dr. med. Christoph Mönninghoff, Dr. med. Martha Dlugaj and Prof. Dr. med. Christian Weimar. This included discussions about medical issues as well as the data they provided for Chapters 2 and 3. In particular, I would like to thank Dr. med. Christoph Mönninghoff for his very important contribution to Chapter 2 of this thesis, for which he provided dozens of manual segmentations, which took up a significant amount of his time. Also, I highly appreciated all the explanations about the medical aspects of the white matter lesions and his help on how to interpret the images. To Dr. med. Martha Dlugaj, thank you for the information about the patients that helped me set up the classification in Chapter 3.

It would not have been possible to keep a proper mental health state if it were not for the "gezelligheid" at the SAS group. A big thanks goes to all my (current and former) colleagues for the nice convivial moments over these four years. To Anneke and Sandra, for the administrative support but especially for their promptness in providing whatever help was needed. To Geert-Jan, not only for the technical support but especially for the endless patience with my data space abuses every now and then. To Bianca, for our many tea breaks that made me forget about my work if only for a couple

of minutes each time; also, thank you for your company in our many trips together, some of them also with Almar. To Almar, an extra thank you for the great Python packages you have developed and that made my work efficiency increase significantly. To Pınar and Jen-Hsuan, for our political/historical/philosophical (not to mention gastronomical) discussions about random topics. To Xiaoying, for always having a word of support. To my two paranymphs, Julia and Meiru, I want to thank for all the cheerful moments we shared, often accompanied by a warm soup, which helped me get through the times of despair and bad weather (which were highly correlated). Also, to the students I have supervised, thank you for having provided me with the experience to teach and yet learn new things (and skills) at the same time. A particular thank you goes to Anne, whose work culminated in Chapter 4 of this thesis; thank you also for your patience with my "niet-zo-goed" Dutch.

To my (other) friends, in the Netherlands but also in Portugal or elsewhere in the world, thank you for the fun and relaxing moments throughout these years, be it in person or via an internet chat. For the sake of not leaving anyone out (and please remember the second sentence of this section), know that if you are reading this, you are most certainly included in the lot.

Finally, to my family: to my parents, for never trusting my pessimism and for having always encouraged me to question things; to my brother and sister, for having taught me how to read, among many other crucial skills; to my nephews, for always being ready to have fun with me, even if this happens more often than not through a computer screen. To all of them, as well as to my brother- and sister-in-law, thank you for your presence, even at the distance. And to Ramon, thank you for all the patience, support and for having always believed in me, especially when I did not.

Rita Simões
Juan-les-Pins, October 2013