

White Matter Lesions and Pattern Recognition in MRI of Neurodegenerative Dementia



Kjetil Oppedal

Faculty of Science and Technology

PhD Thesis UiS no. 313 - November 2016

White Matter Lesions and Pattern Recognition in MRI of Neurodegenerative Dementia

by

Ketil Oppedal

Thesis submitted in fulfillment of the
requirements for the degree of

PHILOSOPHIAE DOCTOR
(PhD)



Faculty of Science and Technology
Department of Electrical Engineering and Computer Science
2016

University of Stavanger
NO-4036 Stavanger
NORWAY
www.uis.no

© Ketil Oppedal, 2016
All rights reserved.

ISBN: 978-82-7644-674-6
ISSN: 1890-1387

PhD Thesis UiS no. 313

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree *Doctor of Philosophy (PhD)* at the University of Stavanger, Norway. The research project has been carried out at the Department of Electrical Engineering and Computer Science, University of Stavanger and Centre for Age-related Medicine (SESAM), Stavanger University Hospital in the period from January 2010 to June 2016. The compulsory courses attended have been given at the University of Stavanger. The work has been funded by The Western Norway Regional Health Authority by grant 911546.

Ketil Oppedal, June 2016

Abstract

Introduction

Expected age is increasing globally and dementia is a common outcome for an increasing number of people. Dementia is a demanding syndrome for the patient and the environment as well as it is costly for society. Damaging changes to the cerebral blood flow also called white matter lesions (WML) are common in the elderly and is expected to increase as age advances. It has been reported that these types of lesions affect cognition in healthy elderly. They are also associated to Alzheimer's disease but have not been much studied in DLB. Quantitative analysis and machine learning have a potential to contribute in understanding the disease process as well as aid in diagnosis.

Methods

Quantitative analysis of WML volumes were calculated using an automatic segmentation routine on magnetic resonance images (MRI) of subjects with Alzheimer's disease (AD), Lewy body dementia (LBD), and normal controls (NC). Statistical tests were performed to compare groups as well as to investigate relations to cognition. Additionally, WML volumes were used as features in a machine learning (ML) environment to check whether WML volume were able to classify subjects with AD and LBD from NC. Texture analysis (TA) may be able to document changes at a microstructural level and was performed in WML and non-WML regions of the different types of MRI's (FLAIR and T1). 2D- and 3D TA features were calculated and used in classification with the aim to serve as a tool for computer aided diagnosis (CAD) in dementia. The dataset used was imbalanced meaning that the number of subjects in each group were very different. Two methods for handling the imbalanced data were tested, namely upsampling and cost-sensitive classification.

Results and conclusions

Severity of WML did neither differ significantly between subjects with dementia and NC nor between mildly demented patients with AD and LBD. WML severity were associated with cognitive decline in AD, but not LBD suggesting that WML contributes to cognitive decline in AD, but not LBD. More studies of the potential clinical impact of WML in patients with LBD are needed.

The best classification results obtained using WML volumes as features in an ML framework discerning subjects with dementia from healthy controls were an area under curve (AUC) of 0.73 and 95% confidence interval of 0.57 to 0.83.

We experienced better classification results when using TA features compared to WML volumes in classification and better results when performing classification on TA features calculated from T1 MRI compared to FLAIR MRI. A total accuracy, reported as mean with standard deviation in brackets over cross validation folds, of 0.97(0.07) or higher was reported for the dementia vs. NC, AD vs. NC, and LBD vs. NC classification problems for both the 2D- and 3D texture analysis approaches. In the AD vs. LBD case a total accuracy of 0.73(0.16) was reported using the 2D TA approach slightly exceeded by the 3D TA approach were 0.79(0.15) was reported.

It seems like the results do not differ much when performing analysis in different regions of the brain and that the results vary in an inconsistent way.

Using upsampling increased classification accuracy to a large extent in the LBD class at the expense of total accuracy and the accuracy of the AD class. In both the two-class problems NC vs. AD and NC vs. LBD, adding cost-sensitivity increased classification performance in many of the tests, but upsampling increased accuracy even more in most of the tests.

High classification performance was achieved when classifying dementia groups from NC's. The classification performance reached when classifying AD from LBD did not reach the same level. Further research with the aim of developing methods with a higher sensitivity to the different brain changes going on in AD and LBD are needed.

Acknowledgments

I want to give my sincerely thanks to my team of supervisors following me through the journey beginning a career in science. All of you welcoming me with warmth and enthusiasm. I remember my main supervisor Prof. Dag Aarsland sitting by my side until late night finishing the application for a PhD grant and especially towards the end of the PhD project, you have shown me trust and given me confidence to finish. I never stop being amazed by how you are able to find time and energy giving guidance to students, collaborating with colleagues, staying very positive and constructive, and simultaneously manage several international research projects.

I remember my co-supervisor Dr. Mona K. Beyer introducing me to the staff at the radiology department – it felt like coming home. I am also grateful for the time we spent together sharing her office space at “Forskningens hus”. Your friendship and supervision has given me a head start. I have been missing having you around after you moved.

I want to mention my co-supervisors Prof. Trygve Eftestøl and Prof. Kjersti Engan together, not because you have contributed less, quite the contrary, but definitely because you work as a team. Through your mutual cooperation you are role models I admire. I will also emphasize my appreciation for the way you have been cheering on me towards the end of my PhD project.

Through methodological discussions, feedback on manuscript writing, thesis writing, and personal encouragement you have all stood by my side from the beginning to the end of this project, through ups and downs even though I several times have only been damp present. Without you I would not even be able to begin this project and at least not finish it.

The project would not be possible without funding from the Western Norway Regional Health Authority.

I find it important to share my appreciation to the working environment at Centre for Age-Related Medicine at Stavanger University Hospital (SESAM), Department of Electrical Engineering and Computer Science at University of Stavanger (IDE), Forskningens hus at Stavanger University Hospital and Department of Radiology at Stavanger University Hospital. I have always been served the possibility of office space, necessary equipment, as well as collegial togetherness.

I want to express my thanks to all the co-authors contributing to the papers enclosed as a part of this thesis.

Without the support from patients, study participants, and volunteers the work related to this thesis would never be able to do.

Thank you for giving access to magnetic resonance images of healthy controls from the ParkWest study.

Eventually I want to give my greatest thanks to friends and family. Lars Rune for countless visits at “Bokkaféen” sharing our very hot custom made Chili Mocca, making me laugh at what I fear and making me relax when I am dispirited or uptight. Prof. Kolbjørn Brønnick for always having your door open. I also want to thank my parents. Most important of all, I want to thank my fiancé and our children. Without your love and commitment I would not stand here today.

Contents

Preface	iii
Abstract	v
Acknowledgments	vii
Contents	xi
1 Introduction	1
1.1 Objectives	1
1.2 Contributions	1
1.3 Dissemination of results	2
1.3.1 Other publications	2
1.4 Organization of thesis	3
2 Medical background	5
2.1 Dementia	5
2.1.1 Alzheimer’s disease	6
2.1.2 The Lewy body dementias	7
2.2 White matter lesions	7
2.2.1 WML and cognitive decline	8
2.3 MRI in dementia	9
2.3.1 Computer aided diagnosis	11
3 Material	13
3.1 The DemVest study	14
3.2 Patient inclusion	14
3.3 Dementia diagnosis and clinical assessment	15
3.4 Baseline clinical characteristics	15
3.5 MRI	17
3.5.1 MRI parameters: the DemWest study	19
3.5.2 MRI parameters: the ParkWest study	20
3.5.3 MRI cohort reliability study	20

4	Region of interest extraction	21
4.1	WML segmentation	22
4.2	Segmentation of grey matter (GM), white matter (WM), and normal appearing white matter (NAWM)	23
5	Feature extraction	25
5.1	Regional WML volume	26
5.2	Texture analysis	26
5.2.1	Texture analysis methods	27
5.2.2	Texture analysis in brain MRI in dementia	28
5.3	Local binary pattern texture analysis	29
5.3.1	2D LBP	29
5.3.2	3D LBP	32
5.4	The texture feature vector	36
6	Statistical analysis and machine learning	37
6.1	Statistical analysis	38
6.2	Classification	38
6.2.1	Bayes decision theory and maximum likelihood classification	39
6.2.2	Decision trees and random forest classification	41
6.2.3	Training and testing	44
6.2.4	Performance measures in classification	45
6.3	The imbalanced data problem	48
6.3.1	SMOTE	48
6.3.2	Cost-sensitive classification	49
7	Contributions	51
7.1	Paper I	51
7.1.1	Results for dementia vs. NC	51
7.1.2	Results for AD vs. LBD	52
7.1.3	Results for correlation between WML volume and cognition	52
7.1.4	Results for multivariate regression	55
7.1.5	Discussion	55
7.2	Paper II	59
7.2.1	Results for 1D feature vectors	59
7.2.2	Results for 2D feature vectors	60
7.2.3	Results for 3D feature vectors	60
7.2.4	Discussion	60
7.3	Paper III	63
7.3.1	Results for three class problem, NC vs. AD vs. LBD	63
7.3.2	Results for the two class problem, NC vs. AD+LBD	64
7.3.3	Results for the two class problem, AD vs. LBD	65
7.3.4	Results when using data from Stavanger only	65
7.3.5	Discussion	66
7.4	Paper IV	69

7.4.1	Results - classification of NC, AD, and LBD	70
7.4.2	Results - three two-class problems	70
7.4.3	Results - robustness test	71
7.4.4	Results - correlation of features vs. cognition	71
7.4.5	Discussion	75
7.5	Preliminary results for texture analysis in GM	78
8	Discussion	79
8.1	Summary of contributions	79
8.1.1	Quantitative analysis of WML in dementia	79
8.1.2	Texture analysis in AD and LBD	81
8.1.3	Machine learning in dementia	82
8.1.4	ROI-wise analysis	83
8.2	Limitations	83
8.2.1	Black box system	83
8.2.2	Clinical diagnosis	84
8.2.3	Postmortem brain autopsy	84
8.2.4	Multicenter study	84
8.2.5	WML segmentation	85
8.3	CAD in clinical practice	85
8.4	Future research	85
	Bibliography	87
	Paper I	107
	Paper II	125
	Paper III	131
	Paper IV	149

Chapter 1

Introduction

Global demographic changes result in better living conditions and higher age many places around the world. One consequence is dementia which has been a growing health problem the last few decades. Dementia is a great burden to the diseased subject, the persons in their close relationships as well as to the health care system and society. Advanced computer analysis of brain images has the potential to be a tool aiding in understanding the disease as well as providing diagnosis at an early stage of disease development.

1.1 Objectives

The main objective of this thesis has been to study how the development of white matter lesions (WML) as depicted from T2-weighted FLAIR MR images relate to dementia, especially Lewy body dementia (LBD), as well as using texture analysis and machine learning (ML) techniques to classify groups with different types of dementia and dementia patients from healthy controls at an early stage of disease development. The analysis have been performed on magnetic resonance images (MRI) from the Norwegian DemWest study. Such a tool has the potential to be of support in early detection of disease progression and thereby increase the possibility of engaging the patient in treatment at an early stage of symptom development.

1.2 Contributions

In Paper I, we quantified the total and regional volume of WML in patients with mild Alzheimer's disease (AD), LBD, and healthy controls using an automated segmentation routine and explored the association with cognitive impairment including memory and executive function. In Paper II we used WML volume as well as texture features in an ML framework with the aim of classifying subjects

with dementia from healthy controls. In Paper III and IV, we applied 2D- and 3D texture analysis in white matter (WM), WML regions as well as normal appearing white matter (NAWM) on FLAIR and T1-weighted MR images as a computer based application for dementia diagnosis by constructing an ML system for classification of different types of dementia as well as healthy controls. We also tested different ways of handling the imbalanced data problem by using oversampling and cost sensitive classification.

1.3 Dissemination of results

Paper I: *White Matter Hyperintensities in Mild Lewy Body Dementia*, K. Oppedal, D. Aarsland, M.J. Firbank, H. Sønnesyn, O.B. Tysnes, J.T. O'Brien, M.K. Beyer, *Dement Geriatr Cogn Disord Extra*, 2012.

Paper II: *Using local binary pattern to classify dementia in MRI*, K. Oppedal, K. Engan, D. Aarsland, M. Beyer, O.B. Tysnes, T. Eftestøl, *Proceedings of International Symposium on Biomedical Imaging (ISBI)*, 2012.

Paper III: *Classifying dementia using local binary patterns from different regions in magnetic resonance images*, K. Oppedal, T. Eftestøl, K. Engan, M. Beyer, D. Aarsland, *International Journal of Biomedical Imaging*, 2015.

Paper IV: *Classifying Alzheimer's disease, Lewy body dementia, and normal controls using 3D texture analysis in magnetic resonance images*, K. Oppedal, K. Engan, T. Eftestøl, M. Beyer, D. Aarsland, *Submitted manuscript*, 2015.

1.3.1 Other publications

Paper V: *Multispectral MRI segmentation of age related white matter changes using a cascade of support vector machines*, S. Damangir, A. Manzouri, K. Oppedal, S. Carlsson, M.J. Firbank, H. Sønnesyn, O.B. Tysnes, J.T. O'Brien, M.K. Beyer, E. Westman, D. Aarsland, L.O. Wahlund, G. Spulber, *J Neurol Sci*, 2012.

Paper VI: *White Matter Hyperintensities and the Course of Depressive Symptoms in Elderly People with Mild Dementia*, Hogne Sønnesyn, Ketil Oppedal, Ole Jacob Greve, Friederike Fritze, Bjørn H. Auestad, Sabine P. Nore, Mona K. Beyer, and Dag Aarsland, *Dement Geriatr Cogn Disord Extra*, 2012, 2(1), 97–111.

Paper VII: *Relationship between orthostatic hypotension and white matter hyperintensity load in older patients with mild dementia*, Hogne Sønnesyn, Dennis W. Nilsen, Ketil Oppedal, Ole Jacob Greve, Mona K. Beyer, and Dag Aarsland, *PLoS One*, 2012, 7(12).

1.4 Organization of thesis

The thesis is divided into eight chapters. Chapter 2 will introduce the reader to the medical background necessary to understand the motivation behind the work presented. The chapter starts with a general introduction to dementia and the two variants of dementia studied followed by an introduction to vascular lesions in the aging brain and their relevance to cognition and dementia. Towards the end of the chapter an introduction to the analyses MRI in dementia will be given ending with a comment on how advanced computational approaches can aid in dementia diagnosis.

In Chapter 3, the data material studied during the work of the thesis will be described. Chapters 4, 5, and 6 takes the reader into more of the technical applications used. The analyses have been performed in several brain regions and Chapter 4 will explain the procedures behind the segmentation of these regions from brain MRI. Both volumetric as well as textural features have been used in this work and Chapter 5 will introduce the reader to these features. Chapter 6 introduces the reader to the classifiers used as well as the classification performance measures. In addition, an introduction to the methods used to handle the challenges experienced when performing data analysis on cohort data will be given.

Chapter 7 presents the reader to the contributions of this thesis by presenting the results obtained during the work of the papers included in this thesis as well as a discussion of these results. Chapter 8 is the last chapter and will give an overall discussion of the thesis project together with some concluding remarks.

Figure 1.1 (below) gives an overview of the workflow throughout the research work presented in the thesis. The figure is divided into four columns each representing a chapter in the thesis. The first column is related to Chapter 3 “Material”, the second column is related to Chapter 4 “Region of interest extraction”, the third column is related to Chapter 5 “Feature extraction”, and the fourth column is related to Chapter 6 “Statistical analysis and machine learning”. The reader will meet the figure again in the beginning of each chapter highlighting the relevant column.

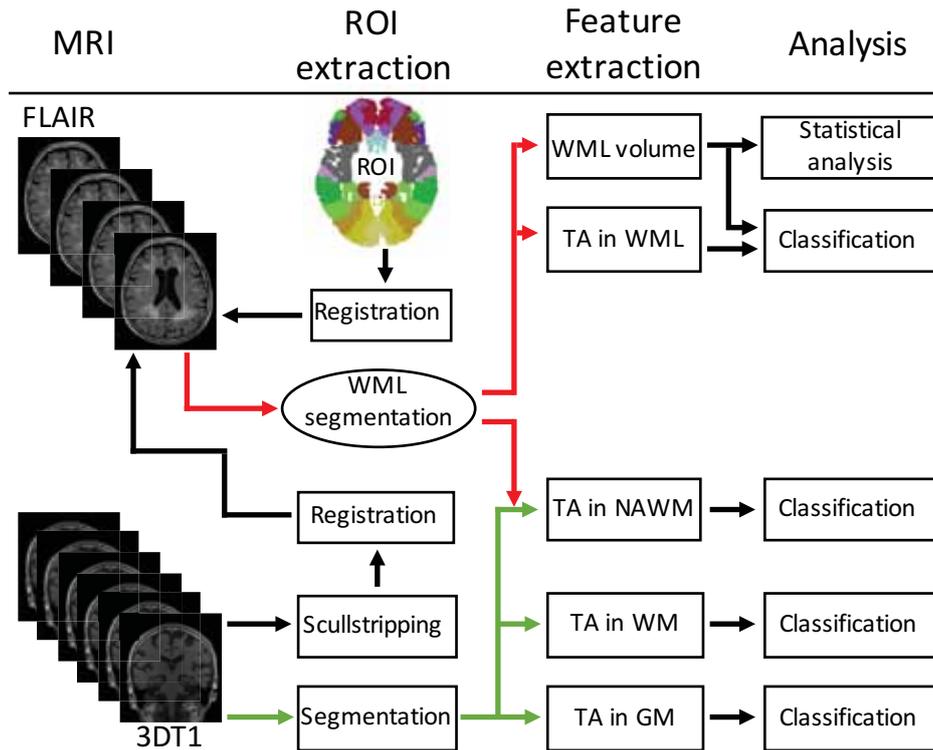


Figure 1.1: The figure gives an overview of the workflow throughout the research work presented in the thesis. The figure is divided into four columns each representing a chapter in the thesis. The first column is related to Chapter 3 “Material”, the second column is related to Chapter 4 “Region of interest extraction”, the third column is related to Chapter 5 “Feature extraction”, and the fourth column is related to Chapter 6 “Statistical analysis and machine learning”.

Chapter 2

Medical background

2.1 Dementia

Cognition are the activities of thinking, understanding, learning, and remembering and can be conscious and unconscious, concrete or abstract, as well as intuitive and conceptual. Cognitive processes use existing knowledge and generate new knowledge. Mental abilities and processes such as knowledge, attention, memory, judgment, evaluation, reasoning, problem solving, decision making, and language develops through the whole lifespan from infancy to the end stages of life. An infant undergo several important cognitive development phases both psychologically and neurologically. During childhood neurons sprout vigorously to form new connections, while in the brain of a teenager, connections between neurons are pruned in a “use it or lose it manner”, as well as myelin insulation is developed improving information processing speed and efficiency. Abstract thoughts, imagining, reasoning ability, goal prioritizing, planning, problem solving, multitasking, the ability to control impulses, understanding of right and wrong, and emotional control, are abilities that evolve well into the twenties. In the same period, brain growth peaks and the development is characterized by maturation. It is important to mention that great variability exists between individuals. At the other end of the life cycle, the brain development takes the opposite direction. As part of healthy aging, typically older adults will experience decreased visual and auditory acuity that interfere with integration of sensory input which may result in slower motor response to sensory stimulation. Other features are loss of recent memory, divided attention, a reduction in overall health status and disruption of formation of new memories.

Some of us will deviate from the path of healthy aging and experience symptoms of neurodegeneration that are part of a dementia syndrome. Typical signs are deterioration in memory, thinking, and behavior in such a way that it inhibits the ability to perform everyday activities. There are many types of dementia, and Alzheimer’s disease (AD) and Lewy body dementia (LBD) are the two most

common neurodegenerative variants. In addition, cerebrovascular disease is another important cause of dementia. Dementia is a progressive syndrome with several stages showing typical symptoms linked to the progression and the degree of neurodegeneration. In the early stage typical experiences are forgetfulness and difficulties keeping track of time and place. During the middle stage forgetting recent events, becoming lost at home, experiencing difficulty with communication, personal care, and behavior changes are common. Eventually, in the last stage of the disease, many experience a total dependency on others with inactivity as a result. Many will be unaware of time and place, have difficulties recognizing people that have been close to you, loose control of muscles and coordination, and changes of behavior will escalate.

Dementia is an umbrella term for a variety of pathological conditions and gives rise to a wide range of symptoms, and the type of dementia is usually classified according to the cause of the experienced symptoms. In clinical practice this is difficult, since the cause can be challenging to reveal. Dementia is defined in various international classification frameworks. The Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) and International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) are most often referred to. Failure of the intellectual functions is a key feature of dementia.

The World Alzheimer Report 2015 [1] states that 46.8 million people worldwide are living with dementia, a number that will almost double every 20 years. These estimates are 12-13% higher than reported in the 2009 edition. The worldwide health care costs related to dementia are calculated as the sum of direct medical costs, direct social care costs (paid and professional home care, and residential and nursing home care) and costs of informal (unpaid) care and are estimated to US\$818 billion. This is an increase in 35.4% compared to the 2010 edition.

The 2014 edition of the yearly updated report states that the strongest causal associations with dementia are those of low education in early life, hypertension in midlife, and smoking and diabetes across the life course. A great challenge in the global society is the increasing age in the population, since age is the primary marker for developing dementia. The importance of dementia research is enormous.

World Health Organization (WHO) states in a fact sheet on dementia from April 2016 (<http://www.who.int/mediacentre/factsheets/fs362/en/>) that 47.4 million people suffers from dementia worldwide, that 7.7 million new cases are expected each year, and that dementia has physical, psychological, social, and economical impact on caregivers, families and society.

2.1.1 Alzheimer's disease

The same fact sheet states that Alzheimer's disease (AD) is the most common neurodegenerative dementia, and accounts for 60-70% of people with dementia.

The most characteristic anatomical findings in AD occurs in both hemispheres (temporal-, parietal-, and frontal lobes), and in areas such as hippocampus, gyrus cinguli, amygdala, nucleus basalis, dorsal raphe, substantia nigra, and locus coeruleus. Hippocampus is always injured and there is reason to believe that the disease starts here. The classical neuropathological signs of AD are amyloid plaques and neurofibrillary tangles [2, 3, 4] as well as granulovacuolar degeneration, deposition of amyloid in the blood vessels, cell damage, and cell death with loss of synapses. The neurofibrillary tangles are formed inside the brain cells and consists of fibers twisted around each other (paired helical filaments) due to hyperphosphorylation of the tau protein. No efficient disease-modifying treatment for AD exists today.

2.1.2 The Lewy body dementias

LBD are the second most common type of degenerative dementia in patients older than 65 years [5]. In [6], the authors state that Dementia with Lewy bodies (DLB) together with dementia associated with Parkinson's disease (PDD) account for 15-20% of people with dementia. In their systemic review, Jones et al. estimated that among persons with dementia, between 0 and 23 % have DLB [7]. The defining pathological feature for these patients is Lewy-body degeneration in brain stem, forebrain, and limbic and cortical structures. The Lewy bodies are abnormal collections of protein called alpha-synuclein developing within the cytoplasm of neurons. They appear as spherical masses that evict other cell components and are eosinophilic cytoplasmic inclusion consisting of a dense body with surrounding radiating fibrils. The DLB and PDD are often combined into a Lewy-body dementia group (LBD) [8, 9]. However, the relationship between localization and density of Lewy-bodies with clinical dementia symptoms is not strong [10], suggesting that other pathologies contribute as well, such as AD pathology, vascular brain changes seen as white matter hyperintensities which are common in the elderly, lacunar infarcts, which may contribute to the clinical presentation of LBD. For example, vascular changes in the basal ganglia are common in the elderly and may cause parkinsonism and cognitive impairment [11]. People with DLB suffer from visual hallucinations, Parkinsonian features such as rigidity and balance disturbances. Patients may experience varying degree of awareness and confusion during the day and many suffer from sleep disturbances.

2.2 White matter lesions

White matter lesions (WML) are frequently observed on brain images of the elderly. In CT images, they are recognized as bilateral, patchy, or diffuse areas with irregular margins of hypodensity, and the hyperintensities on T2-weighted MRI are often used as a surrogate marker. They are typically seen around the ventricles (periventricular WML), but also as focal lesions in the deep white

matter. WML are associated with various disturbances with poor prognosis [12] and can be spotted in periventricular white matter regions, corona radiata, and centrum semiovale. They are reported as a manifestation of cerebral small-vessel disease [13, 14], which is one of the most common of all neurological disorders [15], and they are associated with an increased risk of stroke [16]. European Task Force on Age-Related White Matter Changes was founded in 1996 [17] and LADIS (Leukoaraiosis and disability) was established as an European multicenter collaboration in 2001 [18]. The main study outcome was the transition from an autonomous status to disability, defined as the presence of 2 or more impaired Instrumental Activities of Daily Living (IADL) activities. Secondary outcomes were dementia, stroke, depression, and reduced quality of life [19]. The underlying pathology of WML is heterogeneous, ranging from mild demyelination to incomplete subcortical infarctions, and the exact underlying mechanisms are not fully understood. Wallin and Fladby [20] suggest two mechanisms that could account for the association of WML with dementia. Direct damage to the cortical-subcortical neuronal networks and an interaction between WML and related neuropathological changes, which would imply that the presence of one type of lesion accelerates the expression of the other. In the general population the prevalence of WML ranges from 11-21% in adults aged around 64 to 94% at age 82 [21, 22].

2.2.1 WML and cognitive decline

Understanding the role of WML for the pathogenesis of the progression of cognitive impairment is important, since preventing WML may represent a target for future attempts to prevent or slow down the dementia disease process at an early phase of the disease.

WML becomes more abundant with increasing age in healthy subjects, but they are also found to be associated with dementia [23, 24, 25, 26, 27, 28].

Clinical symptoms associated with WML include gait disturbances [29], depression [25, 30], and cognitive impairment [31]. In [32], the authors studied the effect of normal aging versus hypertension, abnormal body mass index, and diabetes mellitus on WML volume. They found that after the age of 50, the mentioned comorbidities were significantly associated with WML volume.

Cortical changes mediated by WML and vascular risk factors might be associated with cognitive decline and dementia [33]. Mild cognitive impairment, poor episodic memory, and late-life depression are associated with cerebral cortical thinning and WML [34]. Severe WML is associated with worse performances on global tests of cognition, executive functions, speed and motor control attention, naming and visuoconstructional praxis [35]. Increasing severity of WML and number of lacunes (lacunes are 3 to 15 mm cerebrospinal fluid (CSF)-filled cavities in the basal ganglia or white matter, frequently observed coincidentally on imaging in older people, often not clearly associated with discrete neurological symptoms [36]) are both related to worse cognitive performances [37] and

when considered together, WML are significantly associated with cognitive status, whereas the association with lacunes are less prominent. Patients with a combination of severe WML and at least one lacune or of a multilacunar state and moderate-to-severe WML performed more poorly on tests of global cognitive function, psychomotor speed, attention and executive functions, verbal fluency, and working memory compared to those with a combination of less severe changes [38]. Medial temporal lobe (MTL) atrophy, taken as a marker of Alzheimer type pathology, is associated with WML and cognitive functions [39]. In [40], the authors concluded that MTA independently affected memory and language in AD patients and that WML affected attention and frontal executive functions. Together, MTA and WML showed interactions on some cognitive deficits and dementia severity which suggest a combined involvement of Alzheimer and vascular pathology in the earliest stages of cognitive decline. Longitudinal studies show patients with a more severe combination of WML and lacunes present a significantly steeper decline of cognitive performance and a 3-fold risk of developing dementia during follow-up independently of age, sex, education and MTL atrophy [38]. WML severity turns out to be one of the strongest predictor of cognitive decline (dementia and not dementia), independently of age, education, and MTL atrophy. In [41], the authors states that amyloid burden and WML are two common markers of neurodegeneration able to indicate impact on cognition at an early stage in advanced aging. They conclude that amyloid burden and WML had distinct cognitive profiles in a group of clinically normal older adults. The authors found that amyloid burden showed specific influence on episodic memory and that WML were primarily associated with executive function. The findings suggest that both amyloid burden and WML represents neuropathological cascades with distinct etiologies and dissociable influences on cognition even before onset of clinical impairment. In [42], the authors conclude that WML at baseline predicts further development of WML and that its relation to cognitive decline is complex and modulated by brain atrophy. In [43], the authors conclude that white matter hyperintensities contribute to patterns of brain atrophy related to AD. In [44], the authors studied WML detected pathologically postmortem in patients with PDD. They conclude that the individual and cumulative burden of WML, LB lesions, and AD lesions may interdependently contribute to cognitive decline in Lewy body disorders such as PDD. However, the role of WML for the cognitive decline in LBD is not known.

2.3 MRI in dementia

Currently, a definite diagnosis of AD and LBD can only be performed post-mortem. Improved diagnostic techniques may aid in achieving a diagnose earlier and imaging has a potential to add valuable information.

Magnetic resonance imaging (MRI) is an advanced method for non-invasive

construction of soft tissue contrast images. Depending on the chosen MR image sequence, tissue with varying water content provides rise to the image contrast enhancing anatomy or pathology and in even some cases functional activity. The subject under study is positioned in the MRI scanner under the influence of a strong magnetic field, typically 1.5T and 3T, such that the spinning protons precess. The energy from radio waves excites the precessing protons to a higher energy level. Immediately afterwards, relaxation processes gives rise to a nuclear magnetic resonance (NMR) signal that is detected by a receiver coil close to the studied tissue. Field gradients is used to code the NMR signal using frequency- and phase information which makes it possible to locate the NMR signal.

MRI should be used to detect intracranial lesions and rule out other causes to dementia related symptoms [45]. Both DLB and AD are complex diseases with a composite mix of both neurodegenerative and vascular brain changes. Some of these changes can be demarcated on MR images which makes MRI an important tool for studying dementia and cognitive deterioration [46]. The ability to detect neurodegenerative changes early and non-invasively is some of the benefits. Several excellent reviews are available [47, 48, 49]. In [50], the authors reviews available methods for quantitative imaging of white matter anatomy and pathology as well as recent findings in aging and dementia. They state that computer aided quantification offers better statistical power compared to visual rating scales. Early detection of disease and relevant functional connections between brain areas are important benefits. Harper et al. [51] concludes that visual rating scores from MRI offer practical and inexpensive ways of increasing diagnostic accuracy in 184 post-mortem confirmed dementia subjects. They combined several visual rating scores together realizing a higher accuracy.

Some of the MR imaging methods available today provide acceptable anatomical detail as well as being safe for the patient. In the earliest stages of AD and LBD, the characteristic brain abnormalities are not visible on anatomical MRI alone. Quantitative assessment of volumes of the different areas of the brain can be a powerful source of information regarding detection of focal and subtle brain pathology [52, 53, 40, 54]. Giorgio and Stefano provides a thorough review of brain volumetry in clinical applications [55]. Differences in MTL atrophy between AD and healthy controls measured using semi-quantitative techniques on MRI has shown sensitivity and specificity greater than 85% [56, 57, 58]. Significant alterations in tissue microstructure measured by longitudinal and transversal relaxation times was found in patients with DLB compared to healthy controls [59]. Differences between AD and other dementias could not be found with similar sensitivity and specificity though [60, 61, 62]. Higher accuracy may be found measuring hippocampal volume [63]. Cortical thickness measurements have also shown high accuracy discerning AD from healthy subjects [64] and AD from DLB [65, 66]. Watson et al. report that for a similar severity of dementia, DLB appears to have more subcortical atrophy compared to AD [67].

Another MR imaging approach showing promising results concerning analysis of white matter integrity is diffusion tensor MRI (DT-MRI)[68]. DT-MRI is

an MRI method which is sensitive to the diffusion, or Brownian motion, of water molecules. By applying diffusion sensitizing gradients, water molecules will experience a slightly different magnetic field based on a linear relationship between the precession frequency and the position along the direction of the gradient. Under these circumstances will the stationary water molecules precess coherently resulting in no signal loss. On the other hand, the water molecules undergoing diffusion will precess with different phase depending on the length of the displacement along the gradient direction resulting in an attenuated signal. By measuring the water diffusion along at least six diffusion sensitizing directions (preferably more), it is possible to calculate a diffusion tensor using tensor math. Eigen calculations can provide information about the water principal diffusion direction for each image voxel. In [69], Bozzali and Cherubini provides a brief review over DT-MRI analysis as a tool to investigate dementias. On the other hand, Jones et al. [70] points out important limitations to how the information contained in these images is interpreted by many scientists in clinical research.

Measuring changes in the blood oxygenation level-dependent (BOLD) signal, which indirectly quantifies brain activity, is called functional MRI (fMRI) [71, 72]. It has been shown that fMRI is able to differentiate between DLB and AD [73].

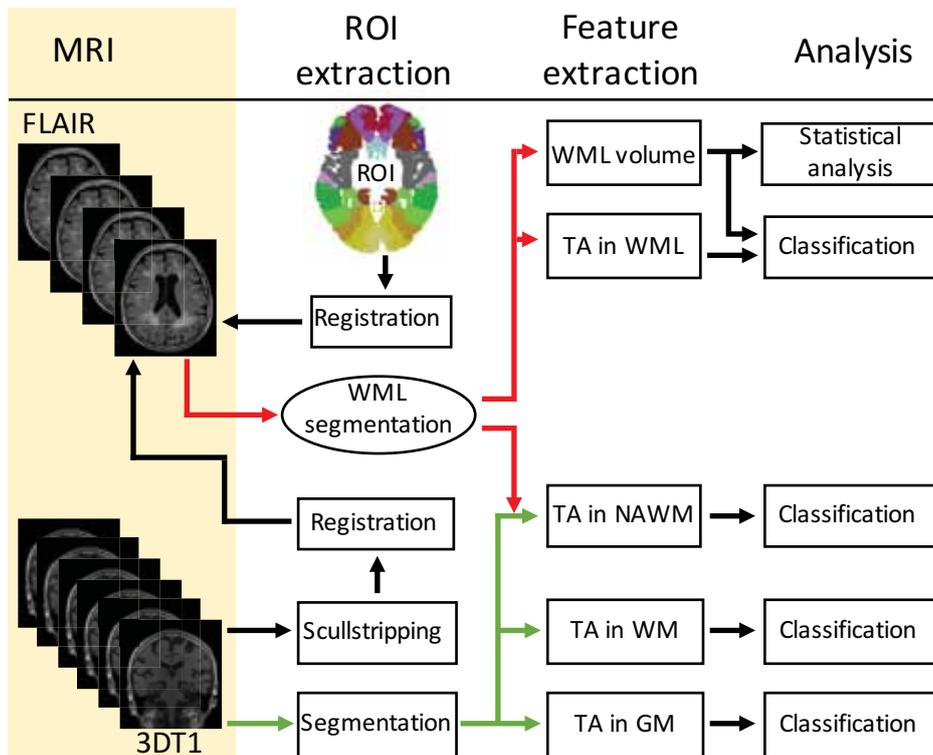
By combining analyses from different variants of MRI as well as combining these with results from positron emission tomography (PET), a multimodal approach has made it possible to track the development of brain changes related to AD as it progresses in time [74, 75]. PET is a functional imaging technique able to assert metabolic processes in the body by measuring the concentration of a positron-emitting tracer delivered to the subject of study.

2.3.1 Computer aided diagnosis

In the context of medical imaging, CAD are computational approaches developed with the ambition to assist medical personnel in the detection of pathology, the quantification of disease progress, and differential diagnosis [76]. Unbiasedness to human mistakes is one benefit using CAD. CAD approaches applied in analyses of neurodegeneration has a great potential when studying dementia [77, 78, 79]. Higher sensitivity and specificity in diagnosing AD from healthy controls have been reported using CAD (95%) as compared to radiologists (between 65 and 95%) [80].

Chapter 3

Material



The data material studied in the work of this thesis is based on subjects with AD and LBD drawn from the DemWest cohort [6] and healthy controls drawn from the ParkWest cohort [81]. In the following chapter the data material will be introduced such that the reader will get an overview of the cohorts, clinical recruitment procedures, relevant baseline clinical information, and MRI protocols.

3.1 The DemVest study

DemWest is an observational study of patients referred for evaluation of cognitive impairment and dementia in all dementia clinics in geriatric, age psychiatric, and neurological departments in western part of Norway. The study started in 2005 and the main objective is to monitor people with various forms of dementia over the course of the disease in order to study the evolution of malfunctions and failures in various brain functions. Problems and symptoms associated with various biological markers of brain disease revealed by magnetic resonance imaging (MRI), spinal fluid examination, genetic testing, and brain examination after death are studied. Particularly interest has been put into DLB.

3.2 Patient inclusion

From March 2005 to March 2007 all referrals to five outpatient clinics in geriatric medicine and old age psychiatry in the counties of Rogaland (Stavanger and Haugesund) and Hordaland (Bergen) in Western Norway were screened. Inclusion criteria were a first time diagnosis of mild dementia according to ICD-10, and a minimum mini mental state examination (MMSE) score of 20. From 2007, patients with DLB and PDD were selectively recruited. Additionally, three neurology outpatient clinics in the same area were contacted, and agreed to refer new dementia cases to one of the participating centres. The patients and their caregivers were first seen by the study clinician, who performed a structured clinical interview of demographic and clinical data. The comprehensive assessment procedure included a detailed history using a semi-structured interview, clinical examination including physical, neurological, psychiatric, and neuropsychological examinations, and routine blood tests. At the time of this study, 235 subjects fulfilled inclusion criteria. Patients are followed annually with the same assessment battery. Patients with acute delirium, terminal illness, recently diagnosed with a major somatic illness, previous bipolar disorder or psychotic disorder were excluded. Further details of the recruitment process are provided elsewhere [6].

A control group consisting of normal controls (NC) was drawn from the Norwegian ParkWest study [81]. They were recruited from spouses, friends, and relatives of patients with Parkinson's disease, and other volunteers, in Western and Southern Norway. These subjects were younger than the dementia popula-

tion, and thus there was a need to individually match for sex, age, and years of education.

The Regional Committee for Medical Research Ethics, Western Norway, approved the study. All participants signed informed consent to participate in the study after the study procedures had been explained in detail to the patient and a caregiver, usually the spouse or offspring.

3.3 Dementia diagnosis and clinical assessment

Diagnoses were made after a detailed assessment, including the use of standardized assessment of hallucinations, parkinsonism and cognitive fluctuations as previously described [6]. The procedures included Hachinski ischemia scale (HIS) [82, 83], Clinical Dementia Rating scale sum of boxes (CDR-SOB) [84], and APOE ϵ 4 genotyping. Transversal 123I-FP-CIT SPECT (DaTscan) images through the basal ganglia were visually assessed for most cases with suspect DLB, scoring caudate and putamen in normal, abnormal, or strong abnormal, by one blinded nuclear medicine physician expert to aid in the diagnosis.

Based on assessments, two old-age psychiatrists independently applied the diagnostic criteria for AD [85] and DLB [10] at baseline and 2 and 5 years later. In cases of disagreement, and in patients fulfilling more than one set of operationalized diagnostic criteria, the final ascertainment was made based on consensus. 46 patients have come to autopsy with a pathological diagnosis consistent with the clinical diagnosis in 85% of the cases. DLB and PDD have similar clinical symptoms and brain changes, and were combined into one group, LBD (12 DLB, 4 PDD) [8, 9]. APOE 4 genotype was analyzed as previously described [86].

A comprehensive clinical assessment battery was applied including standardized cognitive, psychiatric and neurological instruments, as previously described [6]. The neuropsychological battery included tests of verbal memory (California Verbal Learning Test II /CVLT-2) and executive functioning (semantic verbal fluency) in addition to the MMSE. The sum of immediately recalled words from the 5 presentations of the CVLT (list A) and number of animals listed during one minute for the analysis with the WML measures were used. More details are presented elsewhere [6].

3.4 Baseline clinical characteristics

A total of 77 mild dementia subjects, 61 with AD and 16 with LBD, as well as 37 healthy controls had MRI scans of sufficient quality and were included in this study. Please see Figure 3.1 for an overview of patient inclusion. Table 3.1 shows basic clinical characteristics of the subjects from the DemWest cohort that were included in this study and compares them to the subjects that were excluded. Mann-Whitney tests between the two groups showed significantly lower

Hachinski score and the CDR-SOB score in the included subjects compared to the excluded subjects. Age, years of education, sex, MMSE, CVLT-2, verbal fluency, APOE ϵ 4, and CIRS vascular did not show significant differences between groups.

As seen in Table 3.2, there were no significant difference between the AD and the LBD group in age or years of education, but there were more males in the LBD than in the AD group. MMSE was lower and the proportion with at least one ApolipoproteinE 4 (APOE 4) allele (any of the alternative forms of a gene that may occur at a given locus) was higher in the patient groups, as expected, but did not differ between AD and LBD groups. We decided to add APOE 4 as a possible confounder in the statistical analysis, since it is known to be associated with WML [87]. CVLT-2, verbal fluency, CDR-SOB and HIS score did not differ between the two dementia groups, but CIRS vascular was significantly lower in the AD group compared to the LBD group.

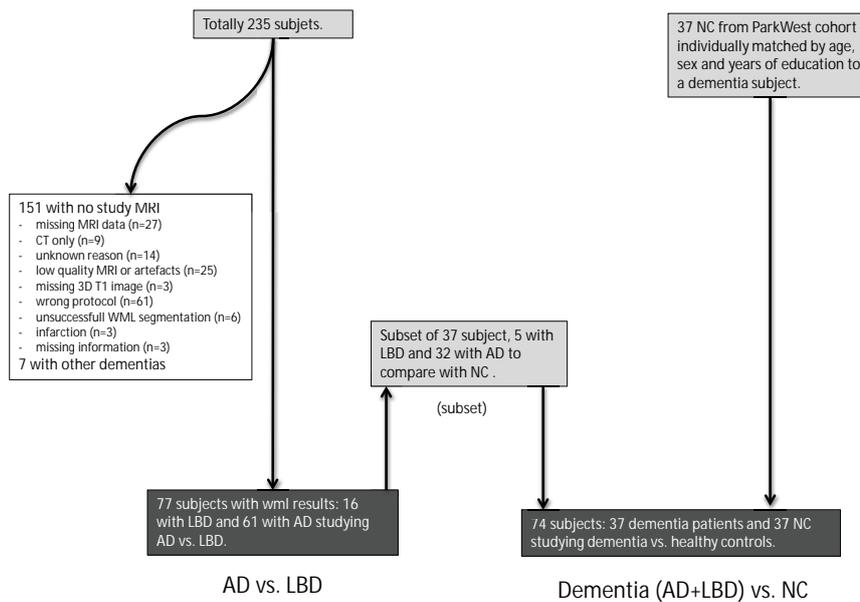


Figure 3.1: Flow chart representing the selection of current MRI study sample from initial DemWest study cohort.

Variable	n(excl) : n(incl)	Excluded	Included	p-value
Age, yr (std)	157:77	75.8 (7.7)	75.2 (7.2)	0.546
Education, yr (std)	145:77	9.7 (3.1)	9.1 (2.4)	0.375
Sex, male:female	158:77	70:87	27:50	0.298 ^{a)}
MMSE (std)	154:77	23.6 (2.8)	24.0 (2.3)	0.387
CVLT-2 (std)	158:75	19.9 (8.4)	19.9 (7.5)	0.849
Verbal fluency (std)	161:76	11.3 (4.4)	11.9 (4.4)	0.640
Hachinski (std)	141:76	4.0 (2.4)	3.0 (2.0)	0.001
APOE ϵ 4, yes:no	100:51	62:38	31:20	0.885 ^{b)}
CDR-SOB (std)	139:74	5.8 (2.7)	4.5 (2.1)	0.001
CIRS vascular	135:73	0.16 (0.51)	0.16 (0.41)	0.477

Table 3.1: Baseline clinical characteristics of the dementia subjects with and without WML volume results. n(excl) = number of patients in the excluded group, n(incl) = number of patients in the included group. Mann-Whitney U test was used except if otherwise indicated. a) Pearson Chi-Square: 2.423, b) Pearson Chi-Square: 0.021, yr= Years, std = Standard deviation, MMSE = Mini Mental State Examination, CVLT-2 = California Verbal Learning Test II, APOE ϵ 4 yes means subject has at least one APOE ϵ 4 allele, CDR-SOB means Clinical Dementia Rating Sum of Boxes Score, and CIRS vascular means Cumulative Illness Rating Scale for vascular symptoms.

3.5 MRI

The patients were scanned at three different sites: Stavanger University Hospital, Haugesund Hospital, and Haraldsplass Deaconess Hospital. A 1.5 T scanner was used in all three centers (Philips Intera in Stavanger and Haugesund, and GE Signa Excite in Bergen), using the same scanner in each center during the entire study period and a common study imaging protocol. The NCs were scanned at four different sites. They were scanned on the same scanners as the patients in Stavanger and Haugesund, and on different scanners in Arendal (1.0T Philips Intera) and Bergen (1.5T Siemens Symphony). The patients and the NC's were scanned on the same scanner using the same protocol at each center.

Two different whole brain MRI sequences were collected for each subject, namely a 3DT1 and a FLAIR MRI. The former provides a high resolution image with good contrast between WM and GM. The latter provides an MR image with a null signal for fluid (e.g. cerebrospinal fluid) and were WML is hyperintense.

After a visual inspection some patients scans were excluded due to either insufficient image quality, not having both FLAIR and T1 images for the patient, or movement artifacts and other artifacts. Please see Figure 3.1 for an overview of inclusion of patients and normal controls.

Variable	NC n=37	AD n=61	LBD n=16	p-value (overall) n=114	p-value AD/LBD n=77
Age, yr(std)	70.6(6.2)	75.2(7.7)	75.5(5.2)	0.005 ^{KW}	0.89
Educ, yr(std)	10.0(2.3)	9.3(2.5)	8.5(1.9)	0.042 ^{KW}	0.176
Sex, m:f	13:24	14:47	13:3	< 0.001 ^{a)}	< 0.001 ^{a)}
MMSE(std)	28.8(1.0)	23.9(2.3)	24.3(2.5)	< 0.001 ^{KW}	0.314
CVLT-2(std), n=61:14	-	19.2(7.0)	23.2(9.3)	-	0.198
Verb flu(std), n=60:16	-	13.0(6.2)	11.5(4.2)	-	0.409
Hachinski(std)	-	2.9(2.1)	3.6(1.9)	-	0.109
APOE ϵ 4, y:n	11:25	27:14	4:6	0.007 ^{b)}	0.133 ^{c)}
CDR-SOB(std)	-	4.4(2.0)	4.8(2.6)	-	0.578
CIRS vasc(std), n = 61:12	-	0.11(0.32)	0.42(0.67)	-	0.045

Table 3.2: Baseline clinical characteristics of the patients and normal controls in the study. Mann-Whitney U test was used except if otherwise indicated. a) Pearson Chi-Square = 18.9, b) Pearson Chi-Square = 9.9, c) Pearson Chi-Square = 2.5, yr= Years, std = Standard deviation, m = male, f = female, y = yes, n = no, Educ = Education, MMSE = Mini Mental State Examination, CVLT-2 = California Verbal Learning Test II, Verb flu = Verbal fluency test APOE ϵ 4 yes means subject has at least one APOE ϵ 4 allele, CDR-SOB = Clinical Dementia Rating Sum of Boxes score, CIRS vasc means Cumulative Illness Rating Scale for vascular symptoms, MW = Mann-Whitney U test, KW = Kruskal-Wallis H test.

3.5.1 MRI parameters: the DemWest study

In Stavanger (Stavanger University Hospital): a *Philips Medical Systems Intera* scanner with software version *NT 10.3.1\PIIM V2.1.4.1 MIMIT MCS* was used. The T1-weighted 3D sequence was a coronal *T1W/3D/FFE*, *Scanning Sequence GR*, and *Sequence Variant SP*, using 10 ms *Repetition Time*, 4.6 ms *Echo Time*, *Flip Angle* 30, 2 *Number of Averages*, 2.0 mm *Slice Thickness* with 1.0 mm *Spacing Between Slices*, *Acquisition Matrix* 256×256 , and an image *Voxel Resolution* (X, Y, Z) of (1.015625 mm, 1.015625 mm, 1 mm) where Z is the slice thickness. The T2-weighted FLAIR image was an axial *T2W/FLAIR*, *Scanning Sequence IR*, and *Sequence Variant SK*, using 6000 ms *Repetition Time*, 100 ms *Echo Time*, 2000 ms *Inversion Time* *Flip Angle* 90, 2 *Number of Averages*, 4.0 mm *Slice Thickness* with 5.0 mm *Spacing Between Slices*, *Acquisition Matrix* 256×202 , and an image *Voxel Resolution* (X, Y, Z) of (0.8984375 mm, 0.8984375 mm, 4 mm) where Z is the slice thickness.

In Haugesund (Haugesund Hospital): a *Philips Medical Systems Intera* scanner with software version *NT 10.3.1\PIIM V2.1.4.1 MIMIT MCS* was used. The T1-weighted 3D sequence was a coronal *T1W/3D/FFE*, *Scanning Sequence GR*, and *Sequence Variant SP*, using 20 ms *Repetition Time*, 16 ms *Echo Time*, *Flip Angle* 30, 1 *Number of Averages*, 1.0 mm *Slice Thickness* with 1.0 mm *Spacing Between Slices*, *Acquisition Matrix* 256×196 , and an image *Voxel Resolution* (X, Y, Z) of (1.015625 mm, 1.015625 mm, 1 mm) where Z is the slice thickness. The T2-weighted FLAIR image was an axial *T2W/FLAIR*, *Scanning Sequence IR*, and *Sequence Variant SK*, using 6000 ms *Repetition Time*, 110 ms *Echo Time*, 2000 ms *Inversion Time* *Flip Angle* 90, 2 *Number of Averages*, 4.0 mm *Slice Thickness* with 5.0 mm *Spacing Between Slices*, *Acquisition Matrix* 256×182 , and an image *Voxel Resolution* (X, Y, Z) of (0.44921875 mm, 0.44921875 mm, 4 mm) where Z is the slice thickness.

In Bergen (Haraldsplass Deaconess University Hospital): a *GE Medical System Signa Excite* scanner with software version *12\LX\MR Software release:12.0_M5_0606.b* was used. The T1-weighted 3D sequence had a *Series Description COR T1 3D FSPGR IR prepped*, *Scanning Sequence GR*, and *Sequence Variant SS\SP\SK*, using 8.224 ms *Repetition Time*, 3.144 ms *Echo Time*, *Flip Angle* 7, 500 ms *Inversion Time*, 1 *Number of Averages*, 1.0 mm *Slice Thickness* with 1.0 mm *Spacing Between Slices*, *Acquisition Matrix* 256×256 , and an image *Voxel Resolution* (X, Y, Z) of (1 mm, 1 mm, 1 mm) where Z is the slice thickness. The T2-weighted FLAIR image had a *Series Description Ax T2 FLAIR*, *Scanning Sequence IR*, using 7927 ms *Repetition Time*, 105.064 ms *Echo Time*, 1981 ms *Inversion Time* *Flip Angle* 90, 1 *Number of Averages*, 4.0 mm *Slice Thickness* with 5.0 mm *Spacing Between Slices*, *Acquisition Matrix* 256×256 , and an image *Voxel Resolution* (X, Y, Z) of (0.8984 mm, 0.8984 mm, 4 mm) where Z is the slice thickness.

3.5.2 MRI parameters: the ParkWest study

In Bergen (Unilabs Røntgen Bergen): a *Siemens Symphony* scanner with software version *syngo MR 2004A 4VA25A* was used. The T1-weighted 3D sequence had a *Series Description t1_mpr_ns_cor*, *Scanning Sequence IR\GR*, and *Sequence Variant SP\MP*, using 2130 ms *Repetition Time*, 3.93 ms *Echo Time*, *Flip Angle* 15, 1100 ms *Inversion Time*, 1 *Number of Averages*, 1.0 mm *Slice Thickness*, *Acquisition Matrix* 256×256 , and an image *Voxel Resolution (X, Y, Z)* of (0.9765625 mm, 0.9765625 mm, 1 mm) where *Z* is the slice thickness. The T2-weighted FLAIR image had a *Series Description t2_tirm_tra_dark-fluid_103*, *Scanning Sequence IR\SE*, and *SK\SP\MP\OSP* using 8400 ms *Repetition Time*, 103 ms *Echo Time*, 2500 ms *Inversion Time* *Flip Angle* 150, 2 *Number of Averages*, 4.0 mm *Slice Thickness* with 4.4 mm *Spacing Between Slices*, *Acquisition Matrix* 256×204 , and an image *Voxel Resolution (X, Y, Z)* of (0.9765625 mm, 0.9765625 mm, 4 mm) where *Z* is the slice thickness.

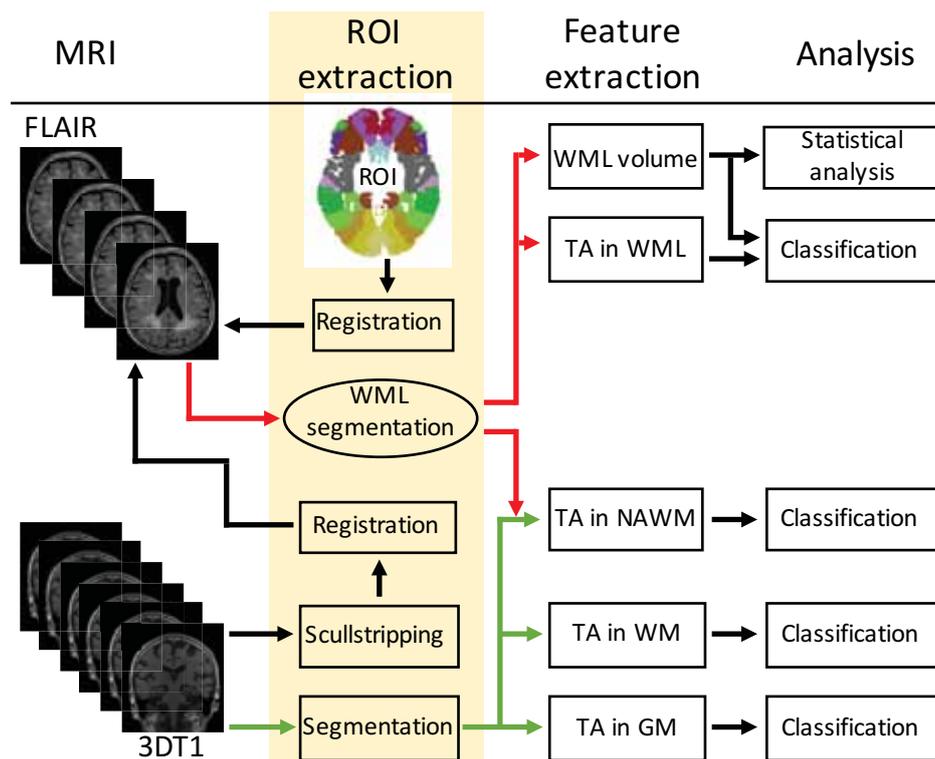
In Arendal: a *Philips Medical Systems Gyroscan NT Inera* scanner with software version *NT 9.1.1\1.3* was used. The T1-weighted 3D sequence was a coronal *T1W/3D/FFE*, *Scanning Sequence GR*, and *Sequence Variant SP*, using 25 ms *Repetition Time*, 6.8912 ms *Echo Time*, *Flip Angle* 30, 1 *Number of Averages*, 2.0 mm *Slice Thickness* with 1.0 mm *Spacing Between Slices*, image *Voxel Resolution (X, Y, Z)* of (1.01562 mm, 1.01562 mm, 2 mm) where *Z* is the slice thickness. The T2-weighted FLAIR image was an axial *T2W/FLAIR*, *Scanning Sequence IR*, and *Sequence Variant SK*, using 6000 ms *Repetition Time*, 100 ms *Echo Time*, 1900 ms *Inversion Time* *Flip Angle* 90, 2 *Number of Averages*, 4.0 mm *Slice Thickness* with 5.0 mm *Spacing Between Slices*, image *Voxel Resolution (X, Y, Z)* of (0.9375 mm, 0.9375 mm, 4 mm) where *Z* is the slice thickness.

3.5.3 MRI cohort reliability study

Since the DemWest cohort is a multicenter study, a reliability study was done where three human volunteers acquired images from all centers using the common cohort protocols. The human volunteers were scanned twice, one hour apart, in all centers to analyze scanner reliability between centers. Scanning was organized with two separate sessions, one hour apart, on the same day. From these scans, the total brain white matter was segmented from the T1 image and the volume calculated for the two MRI acquisitions at all four centers using FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>). Cronbachs alpha between MR scanners at different centers was 0.958. Cronbach's alpha between two time points varied between 0.982 and 0.995, indicating excellent reliabilities both between centers and between different time points. A similar reliability study was performed for the ParkWest study, also with excellent reliabilities.

Chapter 4

Region of interest extraction



Segmentation of WML was mainly performed for three reasons. We wanted to calculate the regional volume of WML in the brain and explore statistical relationships in groups of subjects with AD, LBD, and NC. In addition, we wanted to explore the possibilities that regional WML volume contributes to reduced cognitive score in subjects with AD and LBD. Lastly, we wanted to calculate texture features in WML- and non-WML regions of the brain and do classification discerning subjects with AD, LBD, and NC. The following chapter introduces the reader for the methods used to segment the regions of interest (ROI) used in the work of this thesis.

4.1 WML segmentation

Segmentation of WML was performed according to a method developed and previously published by M J Firbank and colleagues in Newcastle, England [88]. Briefly, the non-brain regions were removed from the T1 image, using the segmentation routines in SPM5 (<http://www.fil.ion.ucl.ac.uk/spm>). After transforming to the image space of the FLAIR image, the segmented T1 image was used as a mask for skull stripping of the FLAIR image. Then the WML were segmented automatically on a slice-by-slice basis from the FLAIR images, with the images in native space, using a threshold determined from the histogram of pixel intensities for each image slice. To explore the regional distribution of WML throughout the brain, a WML region of interest (WML-ROI) template in standard MNI space was used (Montreal Neurological Institute, <http://www.bic.mni.mcgill.ca>). The WML-ROI template was transformed from MNI space to the image space (FLAIR) of each subject by use of the normalization routines in SPM5, and the volumes of WML in each WML-ROI was calculated. The WML-ROI map was based on the Brodmann template, see Figure 4.1.

Because of the variability between the different centers participating in this study, we found it difficult to choose a single threshold level that gave us an acceptable segmentation result without manual editing. A threshold level that gave us an overestimation of the lesion load in every subject was used. Manual editing was then done to correct for this, by removing excess pixels using FSLView (<http://www.fmrib.ox.ac.uk/fsl/index.html>), a medical image-editing program being a part of the FSL software bundle. Manual editing was performed after training by a consultant radiologist who is experienced at evaluation of WML. We performed inter- and intra-rater reliability testing between the two raters to ensure good quality. They both edited the same 10 datasets twice. Once in the beginning to ensure good inter rater reliability and a second time at the end to ensure the similar reliability still persisted and to evaluate intra-rater reliability. We found excellent intra class correlation coefficient (ICC) was 0.998 for inter rater reliability and 0.964 for intra rater reliability.

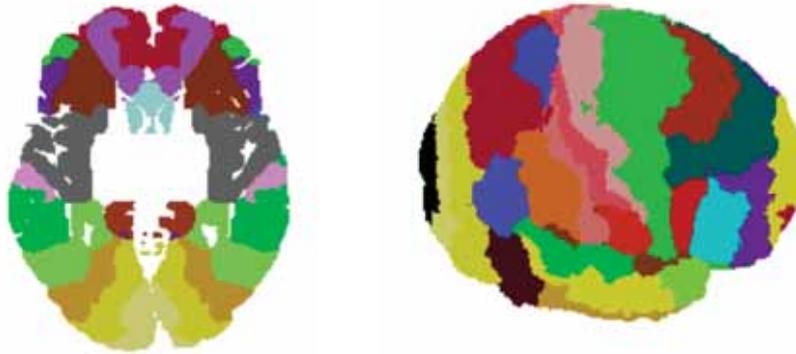


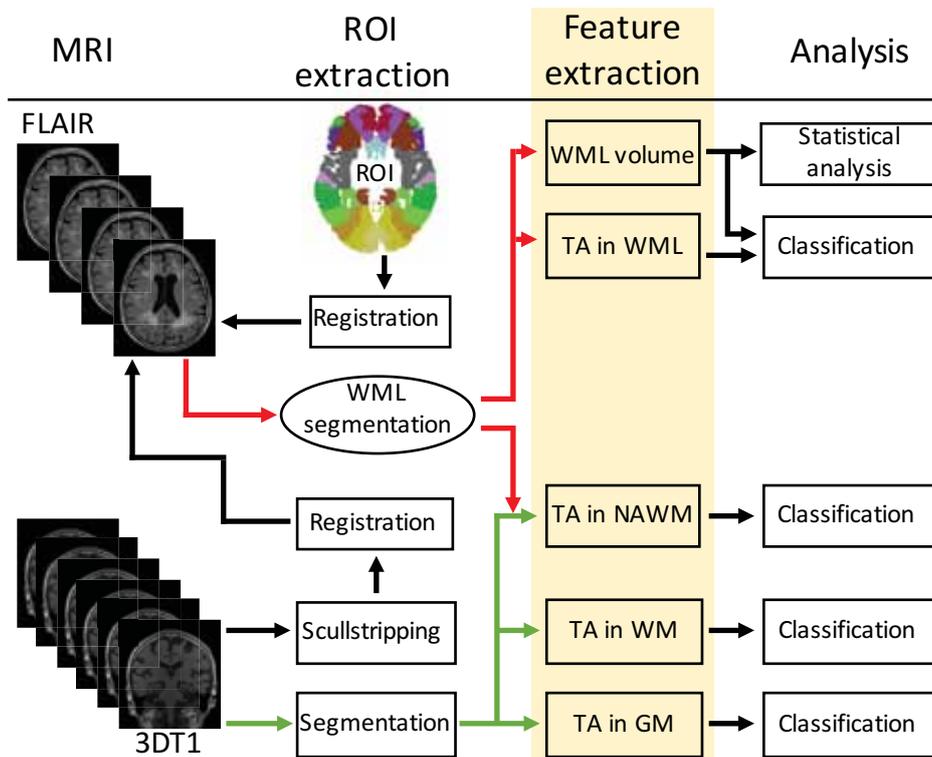
Figure 4.1: Left figure displays an axial slice showing the brodmann areas and the right figure displays a rendered surface image.

4.2 Segmentation of grey matter (GM), white matter (WM), and normal appearing white matter (NAWM)

Segmentation of GM and WM were performed using standard segmentation routines in SPM5 (<http://www.fil.ion.ucl.ac.uk/spm>). The NAWM image was calculated by subtracting the WML image from the WM image.

Chapter 5

Feature extraction



Both volume based features and textural features have been applied. In the first section of this chapter, a description of the volume based features will be given. The next section will provide a general introduction to texture analysis methods as well as an overview of texture analysis used as an application to study dementia. The rest of the chapter is spent introducing the reader to the texture features used in this thesis.

5.1 Regional WML volume

WML volumes in various regions of the brain were used in statistical analyses and as features in classification discerning patients with different types of dementia and healthy controls. From the WML segmentation results the total brain-, total periventricular-, and total deep-white-matter WML volume, as well as the temporal- (Te), occipital- (Oc), frontal- (Fr), parietal- (Pa), cerebellar- (CB), basal ganglia- (BG), central white matter (CW), and corpus callosum (CC) volume in each of the three regions were calculated by counting the number of voxels matching the regions in the Brodmann template and then divided by the total number of voxels in the brain. Total brain volume was obtained calculating the sum of GM and WM. A total of 27 volume features were calculated.

5.2 Texture analysis

In the book “Image processing: dealing with texture” [89] Petrou and Sevilla suggest that “*texture is the variation of data at scales smaller than the scales of interest*”. By this means, texture can be interpreted as the variation that makes for example music and art beautiful, and that makes every day different from the day before.

In optical images, texture may be the result of variation of the albedo or the result of variation of the shape of the imaged surface. In the former case, the change of colors creates variation in the brightness of the image at scales smaller than the scale of the object of interest. In the latter case, texture in an image will arise from the interplay of shadows and better illuminated parts, if the surface is rough, even if it is uniformly colored. In non-optical images as for example MRI, the presence of texture indicates variation of the proton density from one location to another.

Texture properties can be used as descriptors of nature as well as man-made objects, both in two-dimensional and three-dimensional structures. With our senses, texture can be recognized as fine, coarse, smooth, irregular, or lined while touching surfaces [90]. Similar textural properties can be detected visually with new features. Even though our ability to discriminate complex textures visually is limited [91], numerous methods can be applied to quantitatively analyse a vast range of textural parameters that are below the limits of human visual perception [92].

Early examples of quantitative and qualitative use of texture parameters in analysis and classification in industry was when Kaizer used an autocorrelation function to describe aerial photographs in the 1950s [93], and when Haralick used grey tone spatial dependencies as texture features when analyzing photomicrographs, aerial photographs, and multispectral scanner satellite images [90]. These two approaches and other statistical texture analysis methods are used during automatically inspection in machine vision.

In a medical context, texture analysis can apply a quantitative approach to analyse and identify anatomical and pathological structures, and even physiological changes appearing during symptom development. Chien and Fu developed an application in 1974, where co-occurrence matrix was used for automatic analysis of chest X-ray [94]. Application of texture analysis have provided advanced non-visible information in pathological tissue areas in radiograph, ultrasound, computed tomography and magnetic resonance imaging. Harrison [95] extensively reviews the use of texture analysis in a clinical context, analyzing MR images in non-Hodgkin lymphoma, mild traumatic brain injury, and multiple sclerosis. She concludes that “*non visible lesions and physiological changes as well as visible focal lesions of different etiologies could be detected and characterized by texture analysis of routine clinical 1.5 Tesla scans*”.

5.2.1 Texture analysis methods

In their review of texture analysis, Tuceryan and Jain proposed four major categories of texture methods [92], statistical, geometrical, model-based, and signal-processing methods. Other names for analogous methods exist, e.g. Materka and Castellano term the geometrical methods group structural, and the signal-processing methods transform methods [96, 97]. In Kassners and Thornhills review of texture analysis in neurologic MR imaging applications [98], the authors suggest three general categories, namely syntactic, statistical, and spectral methods.

The syntactic texture analysis methods are characterized by a composition of texture elements with certain geometric properties ordered in a pattern by some placement rules. These textures are either analyzed by computing statistical properties from the extracted texture elements, or by extracting the placement rule that describes the texture. Methods comprising syntactic texture is limited in power unless one is dealing with very regular textures and are seldom reported used in texture analysis in neuro-MR imaging.

The statistical methods are based on the spatial distribution of gray values and are one of the early methods proposed in the machine vision literature. The statistical methods are divided into first-order and second-order statistical approaches. Among many first-order variants we find the self explanatory mean of gray level, variance of gray level as well as parameters characterizing local gray-level differences. The second-order statistical features are either extracted from gray level co-occurrence matrixes or run-length matrixes. The co-occurrence

matrix and the run-length matrix were introduced in the 1970s by Galloway and Haralick respectively [99, 90, 100].

As an example of spectral texture analysis methods are features extracted from wavelet functions. Where co-occurrence or run-length features lack the sensitivity to identify larger scale or more coarse changes in spatial frequency, can wavelet functions be designed to evaluate spatial frequencies at multiple scales. Some of the shortcomings of using the Fourier transform for texture description is the lack of ability to delineate temporal changes in frequency content as well as that all signals reflect a superposition of sinusoids. These shortcomings are overcome by the wavelet transform. By trading some degree of spatial-frequency resolution for the ability to localize this frequency content in time, even more flexibility is available.

5.2.2 Texture analysis in brain MRI in dementia

The application of texture analysis in a machine learning (ML) environment has shown success in discerning different dementias from each other and from healthy controls. In [101], Freeborough and Fox used 260 measures derived from the spatial gray-level dependence method, obtained a linear discriminant function using stepwise discriminant analysis obtaining a classification rate of 91% discerning AD from healthy controls. deOliveira et al. [102], found statistical significant differences in gray level co-occurrence matrix measurements in subjects with mild AD, amnesic mild cognitive impairment (aMCI), and healthy controls using Kruskal-Wallis tests and Mann-Whitney U tests. Zhang et al. [103] performed 3D texture analysis of the hippocampus and entorhinal cortex in MR images of Alzheimer's disease. They extracted 3D texture features from image histogram, gradient co-occurrence matrix, and run length matrix and four different feature selection procedures were applied. Depending on ROI selection, feature extraction procedure and selection, they achieved between 64.3% and 96.4% classification accuracy, and that most texture features correlated with the mini-mental state examination (MMSE) score. Sivapriya et al. showed in [104] that texture analysis in brain MRI using wavelets, and classification with back propagation network (BPN) gave high classification accuracy in AD. Li et al. [105], extracted 3D texture features from gray level co-occurrence matrix and run length matrix in the hippocampus area of MR images and found that entropy, grey level non-uniformity, and run length non-uniformity showed significant differences between AD patients, patients with mild cognitive impairment (MCI), and normal controls, and that the texture features were correlated with mini-mental state examination (MMSE) score. This result suggest that 3D texture analysis could describe the pathological changes of hippocampus in patients with early AD and MCI, and be helpful to early diagnosis of AD. Kodama et al. [106], calculated 76 features from co-occurrence matrix and run length matrix in cerebral parenchyma regions and were able to confirm 70.0% of DLB patients, 90.5% of AD patients, and 90.0% of the healthy individuals.

Even though 2D slice by slice approaches are successful, 3D texture features have shown to be an important step towards better discrimination in machine learning systems, especially when the images are intrinsic three dimensional like many MR modalities are [107].

Features from several texture analysis methods can be collected to increase accuracy of classification results. A sophisticated strategy for feature selection with state of the art pattern recognition techniques is required to save computation time, reduce overfitting challenges, and optimize classification results.

5.3 Local binary pattern texture analysis

Ojala *et al.* [108, 109] introduced LBP as a texture operator. Since its discriminative power is high and at the same time computationally simple, LBP is a popular texture descriptor used in various applications, and unifies traditionally divergent statistical and structural models of texture analysis. Adding an image contrast measure (C) calculating the local variance in the pixel neighbourhood, as well as varying the texture neighbourhood enhances the discriminative power of the LBP feature even further. In [110], Unay *et al.* demonstrated that the rotation invariant LBP is invariant to some common MRI artifacts i.e. the bias field.

5.3.1 2D LBP

The derivation of the gray scale and rotation invariant texture operator LBP starts by defining texture T in a local neighbourhood of a monochrome texture image as the joint distribution of the gray levels of $P(P > 1)$ image pixels:

$$T = t(g_c, g_0, \dots, g_{P-1}), \quad (5.1)$$

where gray value g_c corresponds to the gray value of the center pixel of the local neighbourhood and $g_p(p = 0, \dots, P-1)$ corresponds to the gray value of P equally spaced pixels on a circle of radius $R(R > 0)$ that form a circularly symmetric neighbour set. When the coordinates of g_c are $(0, 0)$, the coordinates of g_p are given by $(-R\sin(2\pi p/P), R\cos(2\pi p/P))$ and the gray values of neighbours which do not fall exactly in the center of pixels are estimated by interpolation.

To achieve gray-scale invariance, the gray value of the center pixel (g_c) is subtracted from the gray values of the circular symmetric neighbourhood $g_p(p = 0, \dots, P-1)$, giving:

$$T = t(g_c, g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c). \quad (5.2)$$

By assuming that differences $g_p - g_c$ are independent of g_c and thereby factorizing we get:

$$T \approx t(g_c)t(g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c). \quad (5.3)$$

The distribution $t(g_c)$ describes the overall luminance of the image and is unrelated to local image texture and is removed. The approximated distribution

$$T \approx t(g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c), \quad (5.4)$$

conveys much of the textural characteristics from the original.

By considering just the signs of the differences instead of their exact values, invariance with respect to gray-scale shifts is achieved:

$$T \approx t(s(g_0 - g_c), s(g_1 - g_c), \dots, s(g_{P-1} - g_c)), \quad (5.5)$$

where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0. \end{cases} \quad (5.6)$$

Each sign $s(g_p - g_c)$ is assigned a binomial factor 2^p , such that T is transformed into a unique $LBP_{P,R}$ number that characterizes the spatial structure of the local image texture:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p. \quad (5.7)$$

See also Figure 5.1.

To assign a unique identifier to each rotation invariant local binary pattern, $LBP_{P,R}^i$ is defined as:

$$LBP_{P,R}^i = \min\{\text{ROR}(LBP_{P,R}, i) \mid i = 0, 1, \dots, P-1\} \quad (5.8)$$

where $\text{ROR}(x, i)$ performs a circular bit-wise right shift on the P -bit number x i times.

Certain local binary patterns are fundamental properties of texture. "Uniform" patterns are circular structures that contain very few spatial transitions. They function as templates for micro structures such as bright spot, flat area, dark spot, and edges of varying positive and negative curvature. The uniformity relates to the number of spatial transitions (i.e. bit wise 0/1 changes) in the LBP pattern, e.g. 00000000_2 and 11111111_2 have a uniformity value $U(\text{"pattern"})$ of 0, whereas 00000011_2 and 10000111_2 of 1 and 2 respectively. Patterns that have a U value of at most 2 is designated as "uniform". A gray-scale, rotation invariant, and uniform LBP texture operator is defined as follows:

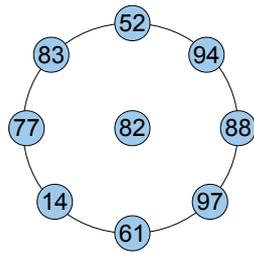
$$LBP_{P,R}^{\text{riu2}} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(LBP_{P,R}) \leq 2 \\ P + 1 & \text{otherwise,} \end{cases} \quad (5.9)$$

where

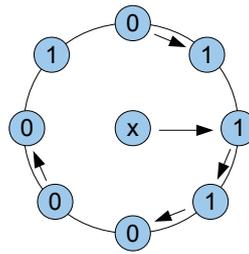
$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)|. \quad (5.10)$$

The value of the LBP code of a pixel (x_c, y_c) is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad s(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$



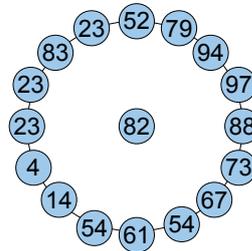
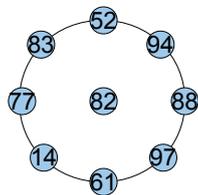
Sample



Threshold

$$x = 1 * 2^0 + 1 * 2^1 + 0 * 2^2 + 0 * 2^3 + 0 * 2^4 + 1 * 2^5 + 0 * 2^6 + 1 * 2^7 = 163$$

Multiply by powers of two and sum



Left figure with small radius and 8 samples.
Right figure with large radius and 16 samples.

Figure 5.1: Top figure demonstrates LBP thresholding. A neighbour to a center pixel is set to one if it has equal or higher pixel value and zero if it has lower. In an anti-clockwise manner every neighbour is multiplied by powers of two and summed as demonstrated in the equation below. The bottom figure demonstrates how the radius and number of samples can be varied in the choice of neighbourhood.

Superscript riu^2 reflects the use of rotation invariant "uniform" patterns that have a U value of at most 2. By definition, exactly $P + 1$ "uniform" binary patterns can occur in a circularly symmetric neighbour set of P pixels whereas the "non-uniform" patterns are grouped under a miscellaneous label ($P + 1$).

The $LBP_{P,R}^{riu}$ and $LBP_{P,R}^{riu^2}$ operators are excellent measures of spatial patterns, but discards contrast. If gray-scale invariance is not required, the contrast (C) of local image texture can be measured with a rotation invariant measure of local variance defined as:

$$VAR_{P,R} = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - \mu)^2, \text{ where } \mu = \frac{1}{P} \sum_{p=0}^{P-1} g_p, \quad (5.11)$$

which is invariant against shifts in gray-scale.

The LBP and C values are calculated for every voxel in the specified region of interest creating an LBP- and a C-valued image. Typically the LBP- and C values are collected and represented as a histogram for each instance in the dataset. The histogram can be used as a vector of features. Other approaches include calculating new features from the histogram.

5.3.2 3D LBP

VLBP

Ojala *et al.* [108, 109] introduced LBP as a texture operator for 2D images. Zhao and Pietikäinen developed an LBP variant for dynamic texture recognition called volume LBP (VLBP) [111] for 3D volumes. The VLBP texture feature takes into account the in-slice neighborhood in a radius R from the pixel of interest sampled with P neighbours as well as the neighborhood in the slice before and the slice after, thus working as a 3D texture operator.

3D texture V in a local neighborhood of a monochrome multislice texture sequence is defined as the joint distribution v of the values of $3P + 2$ given ($P > 1$) image pixels, where P is the number of local neighbouring points around the center pixel in one frame.

$$\begin{aligned} V = v & (s(g_{t_c-L,c} - g_{t_c,c}), s(g_{t_c-L,0} - g_{t_c,c}), \dots, \\ & s(g_{t_c-L,P-1} - g_{t_c,c}), s(g_{t_c,0} - g_{t_c,c}), \dots, \\ & s(g_{t_c,P-1} - g_{t_c,c}), s(g_{t_c+L,0} - g_{t_c,c}), \dots, \\ & s(g_{t_c+L,P-1} - g_{t_c,c}), s(g_{t_c+L,c} - g_{t_c,c})). \end{aligned} \quad (5.12)$$

The gray level value $g_{t_c,c}$ corresponds to the gray level value of the center pixel of the local volume neighborhood. $g_{t_c-L,c}$ and $g_{t_c+L,c}$ correspond to the gray level value of center pixel in the previous and posterior neighbouring frame with distance L . $g_{t,p}$ ($t = t_c - L, t_c, t_c + L; p = 0, \dots, P - 1$) corresponds to the gray level value of P equally spaced pixels on a circle of radius R ($R > 0$) in image t , which form a circularly symmetric neighbour set. As with LBP, invariance

with respect to gray scale is achieved by considering the sign of the differences only $s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$. Each $s(x)$ is assigned a binomial factor 2^q , such that V is transformed from a binary code into a unique $VLBP_{L,P,R}$ number that characterizes the spatial structure of the local image texture:

$$VLBP_{L,P,R} = \sum_{q=0}^{3P+2} v_q 2^q. \quad (5.13)$$

Two variants of a rotation invariant $VLBP_{L,P,R}$ has been developed. The original variant where the rotation invariant LBP is calculated from each frame and then combined and a newer variant obtained by rotating the neighboring set in all three frames synchronously, see [111] for further details as well as Figure 5.2.

LBP-TOP

When increasing the number of samples in the neighborhood, P , the number of VLBP texture patterns increases according to 2^{3P+2} which can get unmanageable large when using large values for P . Another LBP variant for dynamic texture recognition is the LBP three orthogonal planes (LBP-TOP) developed by Zhao and Pietikäinen [111]. Usually, a medical image volume like the anatomical 3DT1 MR image, is thought of as several slices in the XY-direction lined up in a slice-wise manner in the Z-direction. Of course, the alternative where the image is viewed upon as XZ-slices stacked in the Y-direction, or as YZ-slices stacked in the X-direction is equally valid. With this approach the number of LBP patterns is reduced to 3×2^P which is a great advantage. The LBP-TOP patterns are calculated in the XY-, the XZ-, and the YZ-direction of the MR image and are thereby called the LBP_{XY} , the LBP_{XZ} , and the LBP_{YZ} . In the case where MR images are acquired in an isotropic manner, meaning that the in-plane resolution is the same in all three directions, it seems reasonable to set the radius, R , and number of samples, P , equal for all three directions. However, more generally the radii R_X , R_Y , R_Z and the number of samples P_{XY} , P_{XZ} , P_{YZ} can be set differently. The corresponding pattern would then be denoted as $LBP - TOP_{P_{XY}, P_{XZ}, P_{YZ}, R_X, R_Y, R_Z}$ and be an extension of the LBP definition. Suppose the coordinates of the center pixel $g_{t_c, c}$ are (x_c, y_c, z_c) , the coordinates of $g_{XY, p}$, are given by:

$$(x_c - R_X \sin(2\pi p/P_{XY}), y_c + R_Y \cos(2\pi p/P_{XY}), t_c), \quad (5.14)$$

the coordinates of $g_{XZ, p}$, are given by:

$$(x_c - R_X \sin(2\pi p/P_{XZ}), y_c, t_c - R_Z \cos(2\pi p/P_{XZ})), \quad (5.15)$$

and the coordinates of $g_{YZ, p}$ are given by:

$$(x_c, y_c - R_Y \cos(2\pi p/P_{YZ}), t_c - R_Z \sin(2\pi p/P_{YZ})). \quad (5.16)$$

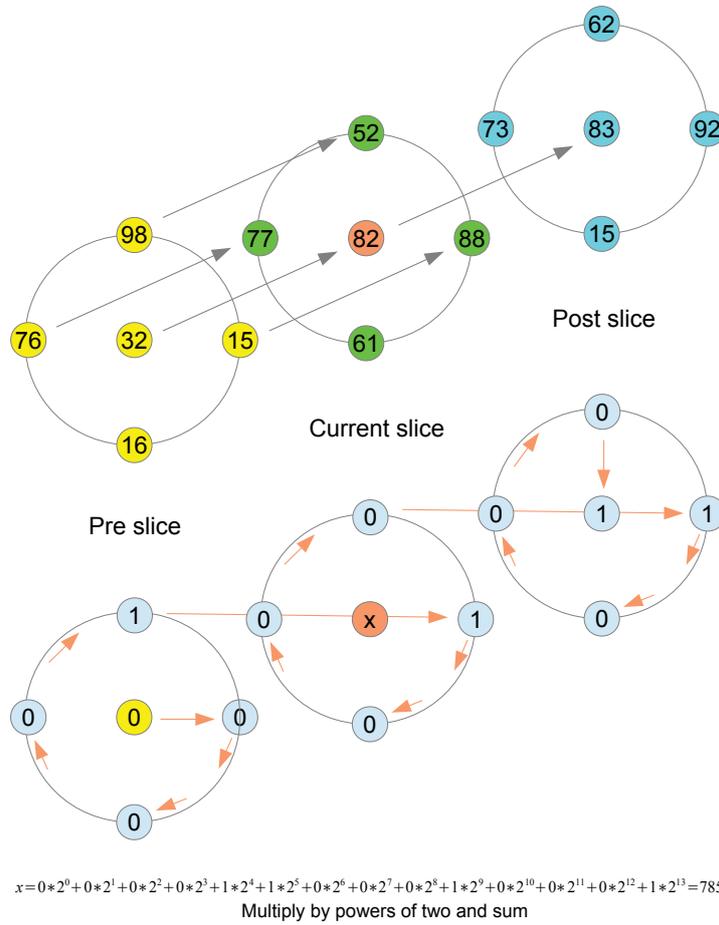


Figure 5.2: Top figure demonstrates the neighborhood configuration in VLBP. The center pixel in the current slice (orange), named x , is the current pixel. A threshold is set on each pixel in the 3D neighbourhood. The bottom figure shows how the readout is performed, which starts at the center pixel in the pre slice (yellow) and ends in the center pixel in the post slice. Every neighbour is multiplied by powers of two and summed as demonstrated in the equation below.

Contrast

The $VLBP_{P,R}^i$ and $LBP - TOP_{P,R}^i$ operators are excellent measure of the spatial pattern, but discards contrast. If gray-scale invariance is not required, the contrast (C) of local image texture can be calculated as local variance and is invariant to rotation and shifts in gray-scale. In the VLBP case it will be defined as:

$$C_{L,P,R} = \frac{1}{3P+2} \sum_{p=0}^{3P+1} (g_p - \mu)^2, \quad (5.17)$$

$$\text{where } \mu = \frac{1}{3P+2} \sum_{p=0}^{3P+1} g_p,$$

and in the LBP-TOP case:

$$C_{P,R} = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - \mu)^2, \text{ where } \mu = \frac{1}{P} \sum_{p=0}^{P-1} g_p, \quad (5.18)$$

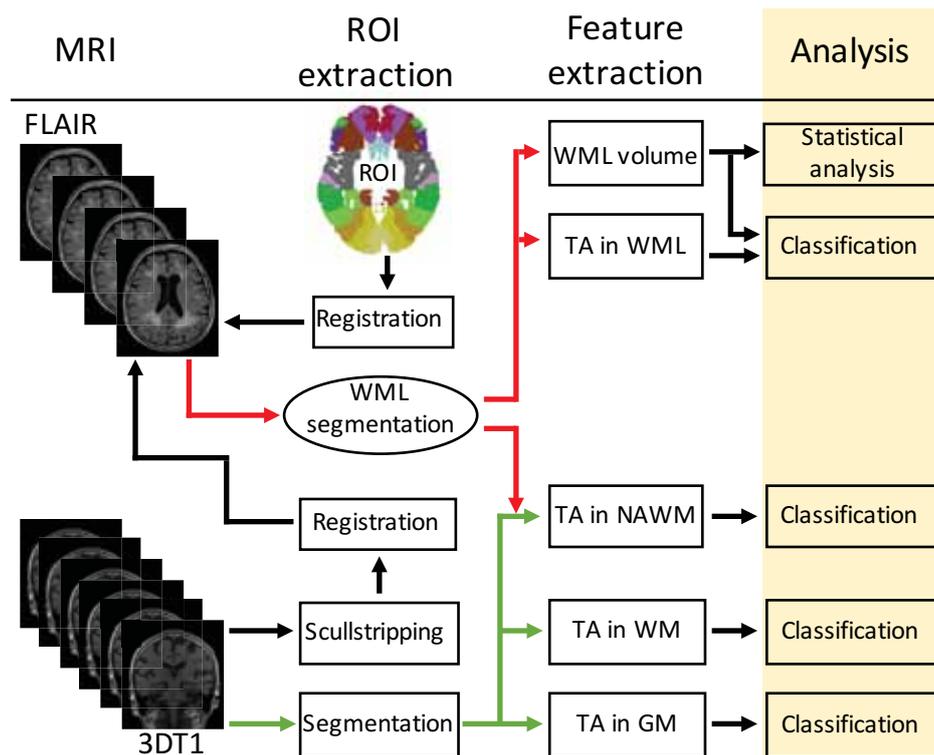
calculated in all three directions (XY, XZ, and YZ).

5.4 The texture feature vector

The local binary pattern approach to texture analysis produces a vast amount of information that can be interpreted in many ways producing a range of different results. Very common is the calculation of an LBP histogram. Especially when used with high sampling of the neighborhood, as when using a 3D variant of LBP, the histograms grows large. Large feature vectors in combination with few subjects in the training dataset introduces limitations to classification and poor accuracy due to overtraining. Instead of using these histograms or the whole range of LBP values, we decided to calculate statistical measures of the collection of LBP values and use the generated numbers as features in classification. The calculated measures were mean, standard deviation, variation, median, interquartile range, entropy, skewness, and kurtosis of the LBP values. For each neighborhood configuration eight features related to LBP and eight related to C were calculated and concatenated into one feature vector for each subject.

Chapter 6

Statistical analysis and machine learning



Using the quantified WML volumes and calculated texture features an important objective was to compare different dementia groups and healthy controls with statistical tests and discern the same groups using classification and machine learning techniques. In the following chapter the reader will be presented for the statistical tests as well as the classifiers and other machine learning techniques used in the work of this thesis.

6.1 Statistical analysis

Comparison of demographic variables was done with the nonparametric Mann-Whitney and Kruskal-Wallis tests because of the non-normal distribution of the data for age, years of education and MMSE score. Neither of the above mentioned features nor data for WML volumes became more normally distributed after a log transformation. For sex a Pearson Chi-Square was used to test independence between groups. We first compared the matched dementia and NC groups, and subsequently the AD vs. LBD group. WML volume measurements were compared between the dementia group vs. normal controls and between the AD group and LBD with the nonparametric Mann-Whitney test. Correlations between WML load and cognitive scores were done calculating the Spearman correlation coefficient with two-tailed significance. We also applied Hierarchical multivariate linear regression analysis to study the associations between cognitive scores and WML volume, adjusting for demographic variables significantly associated with cognition in bivariate analysis. All statistical tests were performed using PASW Statistics 18, release 18.0.1, and p-values < 0.05 was considered statistically significant.

6.2 Classification

Classification of objects into a number of categories is a scientific discipline named pattern recognition [112]. Depending on the application, these objects can be images or signal waveforms or any type of measurement that need to be classified. Machine learning evolved from pattern recognition and was defined by Arthur Samuel in 1959 as “the field of study that gives computers the ability to learn without being explicitly programmed”. Another way of putting it is that machine learning is about algorithms that can learn from data and perform predictions from the learned algorithms.

Pattern recognition systems typically consist of five elements: sensing, segmentation and grouping, feature extraction, classification and post processing. A camera or a microphone array could be examples of transducers or sensors that typically can be input to a pattern recognition system. Characteristics and limitations of the transducer as for example the bandwidth, the resolution, the sensitivity, the distortion, the signal-to-noise ratio, or the latency induces difficulty into the equation. In this thesis MR images formed by radio frequency

(RF) pulses applied to a subjects brain inside strong magnetic fields (1.5T) is the source of input to the analysis system.

Many classification algorithms exists, and it is common to divide them into two main subgroups, unsupervised and supervised techniques. The former refers to algorithms where for example a function is inferred with the purpose of unveiling hidden structures in unlabeled data. The latter refers to algorithms where a function is inferred on the background of labeled data. The labeled data have known class affiliations. Using this information, a classifier is trained based on a subset of the available data (the training data) and afterwards validated using the remaining data (the test data).

6.2.1 Bayes decision theory and maximum likelihood classification

Many pattern classification methods exists and there is a vast amount of literature in the field. Bayesian decision theory is a fundamental statistical approach for many classification methods. It is based on an assumption that the decision problem is posed in probabilistic terms. Given a problem with j classes ω_j for $j = 1, 2, \dots, c$ with prior probability $P(\omega_j)$ and feature vector \mathbf{x} in a d -dimensional Euclidean space \mathbf{R}^d called the *feature space* – a set of random variable whose distribution depends on the class condition – the class-conditional probability density function is expressed as $p(\mathbf{x}|\omega_j)$. The joint probability density of finding a pattern that is in category ω_j and has feature vector \mathbf{x} can be written in two ways: $p(\omega_j, \mathbf{x}) = P(\omega_j|\mathbf{x})p(\mathbf{x}) = p(\mathbf{x}|\omega_j)P(\omega_j)$. Rearranging the equation leads us to *Bayes formula*:

$$P(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{p(\mathbf{x})}, \quad (6.1)$$

where

$$p(\mathbf{x}) = \sum_{j=1}^c p(\mathbf{x}|\omega_j)P(\omega_j), \quad (6.2)$$

and can informally be expressed as

$$posterior = \frac{likelihood \times prior}{evidence}. \quad (6.3)$$

The *a posteriori* probability $P(\omega_j|\mathbf{x})$, namely the probability of the class being ω_j given the values of feature vector \mathbf{x} , can thus be calculated from the prior probability $P(\omega_j)$ and the *likelihood* $p(\mathbf{x}|\omega_j)$ of ω_j with respect to \mathbf{x} where $p(\mathbf{x})$, called the *evidence* factor, can be viewed as a scale factor that ensures that the posterior probabilities sum to one. In a classification problem a given \mathbf{x} will belong to the class ω_j showing the highest posterior probability $P(\omega_j|\mathbf{x})$. This decision minimizes the probability of misclassification.

In pattern recognition applications, we seldom have enough knowledge about the complete probabilistic nature of the problem at hand. We may only have a

set of training data which represents the patterns we want to classify, and by using this information we can estimate the unknown probabilities and probability densities and use them as if they were the true values. Estimating the prior probabilities presents no difficulty, but the class-conditional densities can be challenging to determine, and the number of training samples easily seems too small, especially if the feature vector \mathbf{x} is large. By assuming that the class-conditional densities $p(\mathbf{x}|\omega_j)$ are normal distributed with mean μ_j and covariance matrix Σ_j the problem is simply one of estimating these parameters instead of the probability function. Parameter estimation is a classical problem in statistics and can be approached in many ways. One of the main benefits of using the Maximum-likelihood estimation methods is its good convergence properties as the number of training samples increases. The parameters are seen upon as quantities whose values are fixed but unknown, and that the best estimate is defined to be the one that maximizes the probability of obtaining the samples actually observed.

Given a set of samples separated in classes such that we have c datasets, D_1, \dots, D_c and the samples in D_j drawn independently according to the probability law $p(\mathbf{x}|\omega_j)$. Let us further assume that $p(\mathbf{x}|\omega_j)$ is uniquely determined by the value of a parameter vector θ_j which for a normal distributed sample set $p(\mathbf{x}|\omega_j) \sim N(\mu_j, \Sigma_j)$, consists of the components μ_j and Σ_j . Explicitly, the dependence of $p(\mathbf{x}|\omega_j)$ on θ_j , is written as $p(\mathbf{x}|\omega_j, \theta_j)$. Using the information provided by the training samples, our problem is to obtain good estimates for the unknown parameter vectors $\theta_1, \dots, \theta_c$ associated with each category. We assume that the parameters for the different classes are functionally independent. D_j is a set of training samples drawn independently from the probability density $p(\mathbf{x}^{(j)}|\theta_j)$ and contains n samples, $\mathbf{x}_1^{(j)}, \dots, \mathbf{x}_n^{(j)}$.

We then have $p(D_j|\theta_j) = \prod_{i=1}^n p(\mathbf{x}_i^{(j)}|\theta_j)$. $p(D_j|\theta_j)$ is called the *likelihood* of θ_j with respect to the set of samples in class j also written $l(\theta_j|\mathbf{x}_i^{(j)})$. The *maximum-likelihood estimate* of θ_j is the value $\hat{\theta}_j$ that maximizes $l(\theta_j|\mathbf{x}_i^{(j)})$ written as:

$$\hat{\theta}_j = \arg \max_{\theta} l(\theta_j|\mathbf{x}_i^{(j)}) \quad (6.4)$$

By applying standard methods of differential calculus the global maximum can be estimated and it can be shown that the solution is the sample mean $\mu_j = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^{(j)}$ and the weighted average of the sample covariance matrices $\Sigma_j = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i^{(j)} - \mu_j)(\mathbf{x}_i^{(j)} - \mu_j)^T$ for each class ω_j .

Based on the means and covariance matrices estimated from the training samples $\mathbf{x}_1, \dots, \mathbf{x}_n$ drawn from D_j , discrimination functions for each class ω_j can be produced and a sample earlier unseen \mathbf{x}^{test} can be classified as belonging to class ω_1 if $p(\mathbf{x}|\omega_1)P(\omega_1) > p(\mathbf{x}|\omega_2)P(\omega_2)$ and otherwise decide ω_2 in a two-class problem.

6.2.2 Decision trees and random forest classification

In many situations our classification problem involves nominal data i.e. discrete descriptors without natural notion of similarity or ordering. In these situations a typical feature vector will be a list of attributes and not real numbers. To consider such discrete problems, rule-based or syntactic pattern recognition methods are applied.

An example of such a method is a decision tree, which is created by asking a sequence of questions concerning the values in the feature vector in an orderly fashion. The first value considered makes the root node at level zero in the decision tree construction, and is thereby displayed at the top. Successive links or branches from the root node to descendant nodes in the tree are formed based on answers from questions, for example a "yes/no" question or "true/false" question or "value(property) \in set-of-values" question. These are similarly connected until a terminal or leaf node is reached which has no further branches. The test pattern is assigned the category of the leaf node.

Before classification can be performed, the decision tree has to be created using training data. Using a set D containing instances having a class label and a set of discriminative properties, the decision tree progressively splits the set into smaller and smaller subsets. If the new subset is pure, i.e. all the samples in the subset have the same category label, this part of the tree can be terminated and the node considered a leaf node. If on the other hand the subset had a mixture of labels, a decision whether to stop and accept an imperfect decision, or select another property and grow the tree further, is appropriate.

This methodology provides us with a recursive tree-growing process and gives rise to a generic tree-growing procedure called CART (classification and regression trees), and provides a general framework that can be instantiated in various ways to produce different decision trees. Important design decisions of the CART methodology are how many splits will be performed at each node, which properties should be tested at each node, when should a node be declared a leaf, how to prevent the tree growing excessively large, how to assign category labels if the node is impure, and how to handle missing data. Many of these design decisions can be difficult to handle.

Some of the experienced benefits using decision trees are the ability to handle large datasets using standard computing resources, the boolean logic making the system highly interpretable, and the ability to handle both numerical and categorical data. One drawback is related to the search methods used when looking for the optimal split. For example using the greedy algorithm may result in decisions made on local optima. Another drawback is the creation of overly complex trees resulting in overfitting, reduced generalization ability, and thereby poor validation performance. A common approach to reduce the limitation introduced by overfitting is pruning which means that sections providing little classification power are removed with the aim of reducing the size of the tree without reducing classification performance. A third drawback could be

problems concerning data with many input variables each one containing small amount of information. In extreme cases, a single tree classifier will have only slightly better performance compared to a random choice of class. Many of the aforementioned challenges are overcome by the random forests algorithm which was developed by Leo Breiman [113].

One of the key factors behind the success of Random Forest classifiers is the introduction of randomness to the tree growing process which have to important layers: the use of bagging and random features in the split selection. Together, they improve strength as well as reducing the correlation between each tree in the forest.

Bagging means that for the chosen training set a sub-sample is drawn with replacement from the original training set. A classification tree is grown on the new sub-sampled training set using random feature selection. Given a dataset S containing several instances $1, 2, \dots, N$, each instance consisting of several descriptors a, b, c and a class label C , see Equation 6.5. S_1, S_2, \dots, S_M , are subsets of S , $S_1 \subseteq S$, $S_2 \subseteq S$, \dots , $S_M \subseteq S$, sampled with replacement, see Equation 6.6 and the top part of Figure 6.1.

$$S = \begin{bmatrix} f_{a1} & f_{b1} & f_{c1} & C_1 \\ f_{a2} & f_{b2} & f_{c2} & C_2 \\ \vdots & \vdots & \vdots & \vdots \\ f_{aN} & f_{bN} & f_{cN} & C_N \end{bmatrix} \quad (6.5)$$

$$S_1 = \begin{bmatrix} f_{a1} & f_{b1} & f_{c1} & C_1 \\ f_{a7} & f_{b7} & f_{c7} & C_7 \\ \vdots & \vdots & \vdots & \vdots \\ f_{a23} & f_{b23} & f_{c23} & C_{23} \end{bmatrix}, S_M = \begin{bmatrix} f_{a2} & f_{b2} & f_{c2} & C_2 \\ f_{a12} & f_{b12} & f_{c12} & C_{12} \\ \vdots & \vdots & \vdots & \vdots \\ f_{a35} & f_{b35} & f_{c35} & C_{35} \end{bmatrix} \quad (6.6)$$

For each subset a classification tree is generated. A forest of M classification trees is build from the M random subsampled datasets, see lower part of Figure 6.1. During the tree growing process, the nodes in the classification trees are split using random features. That means that only a subset of all the features are used when growing the tree at each node. Finally, a new instance that is run through the classification tree will be classified according to the mode of the decisions from the trees in the forest. Bagging introduces two benefits to the classifier development. One, accuracy is enhanced when used together with random feature selection. Two, bagging can be used to give ongoing estimates of the generalization error of the combined ensemble of trees, as well as estimates for the strength of and correlation between each tree based on the non-sampled instances from the original training set. It can be shown that the Random Forest method is robust to overfitting as more trees are added to the classifier [113].

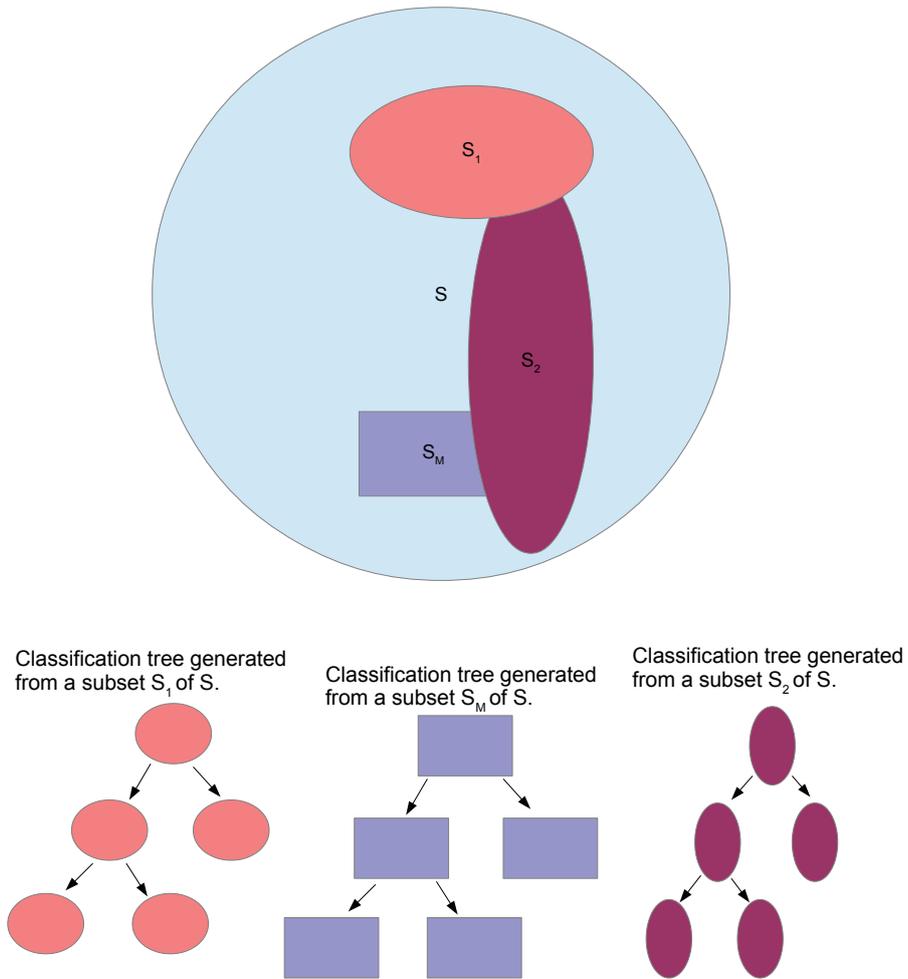


Figure 6.1: A dataset S and subsets S_1, S_2, \dots, S_M sampled with replacement from S . A classification tree is generated for each subset S_1, S_2, \dots, S_M . The final decision is made based on the majority vote of classification results from the forest.



Figure 6.2: 5 folds cross validation using 4/5 of the data for training a classifier and 1/5 of the data to test the classifier. The process is repeated five times, each time using a different fold for validation. The resulting validation measure is averaged over all five folds.

6.2.3 Training and testing

In a medical context an important application of pattern recognition is computer-aided diagnosis. The final diagnose decision is made by a medical doctor. Therefore, an aim of pattern recognition applications could be to assist the doctor while interpreting medical image information. Poor image quality can alter the subtle nature of the findings and subjectivity can prevent accurate results. A common challenge using machine learning techniques such as classification of subjects having a disease from healthy controls is the lack of sufficient amount of data. A consequence may be difficulty in learning a classifier with sufficient low bias and variance in the validation results. It is of utmost importance that great care is taken when training a classifier on data containing few instances and a high number of descriptors for each instance. When performing training under such circumstances, overfitting is a common problem and needs to be dealt with. In addition, the performance measures calculated from the test data can become unreliable showing high bias and variance. Overfitting refers to a learned classifier that performs well on the training data, but have low generalization abilities resulting in poor validation outcomes.

A widely used method when learning a classifier on scarce data is cross-validation (CV). When doing K-fold CV, data are divided into K roughly equal-sized partitions, see Figure 6.2. If K equals five, training is performed on four fifths of the data and testing is performed on the remaining one fifth. The process is repeated K times and the validation error is averaged over all K folds.

A special case of K-folds CV is when K equals the number of instances in the dataset such that the validation is performed on one instance only. This method, called leave-one-out CV (LOOCV), makes the most out of the data concerning

training and another benefit is that the validation bias is commonly low. Two drawbacks are the high validation variance [114] and that the total prediction error becomes overly optimistic [115]. Five- or tenfold CV is recommended as a compromise between validation bias and variance [116, 117]. Even better performance can be achieved when using stratification. That means that the relative frequency of instances in each class is kept the same throughout the CV procedure.

In Hastie et al. [118], the authors mention important elements regarding CV and warn against using the methodology wrongly. Typical misuse is when LOOCV is used for training on sparse data with high number of descriptors. In [115] they achieved 84% apparent predictive ability on a synthetic dataset consisting of random numbers only. Two other wrong approaches are when the complete dataset has been used for feature sub selection ahead of using CV for classifier learning, parameter optimization and validation, or when classifier learning is performed using CV and then validation is performed on all data. A third option is when the result from validation using CV is used for optimizing parameters during classifier learning. Rao et al. [119] emphasize the dangers when using results from CV to choose the most optimized classifier from a too large set of possibilities. In all the above mentioned cases the learned classifier is not validated on true unseen data. Taylor et al. [120] attaches great importance to the need for validation on unseen data, or external validation. In the case of sparse data with many descriptors, descriptor (or feature) sub selection is an important step when training a classifier [121], but it is important that this step is included in the CV scheme and thus repeated for every fold. Two common approaches for feature subset selection is the filter approach and the wrapper approach [122]. An example of the former approach is when the data is filtered such that the features that correlates highly with the class labels and at the same time little with each other are collected. The latter approach uses classification results to test different feature subsets.

Hjorth et al. [123] was first out using one CV procedure inside another. In [115] Anderssen et al. points to the importance of using an inner CV procedure for feature selection and classifier training and an outer CV procedure for validation, see Figure 6.3 to ensure that the developed model is validated on unseen data. This is called nested CV and are verified in [124, 125] as well.

6.2.4 Performance measures in classification

Traditionally, *accuracy* and *error rate* have been widely used as performance measures in classification. *Accuracy* is calculated as the sum of correct classified instances in all classes divided by the total instances in the dataset and *error rate* is defined as: $error\ rate = 1 - accuracy$. These measures are frequently used, but can be very sensitive to the composition of the dataset. For example, if the relative number of instances in each class is very different the results can be deceiving not providing adequate information [126, 127, 128, 129, 130, 131,

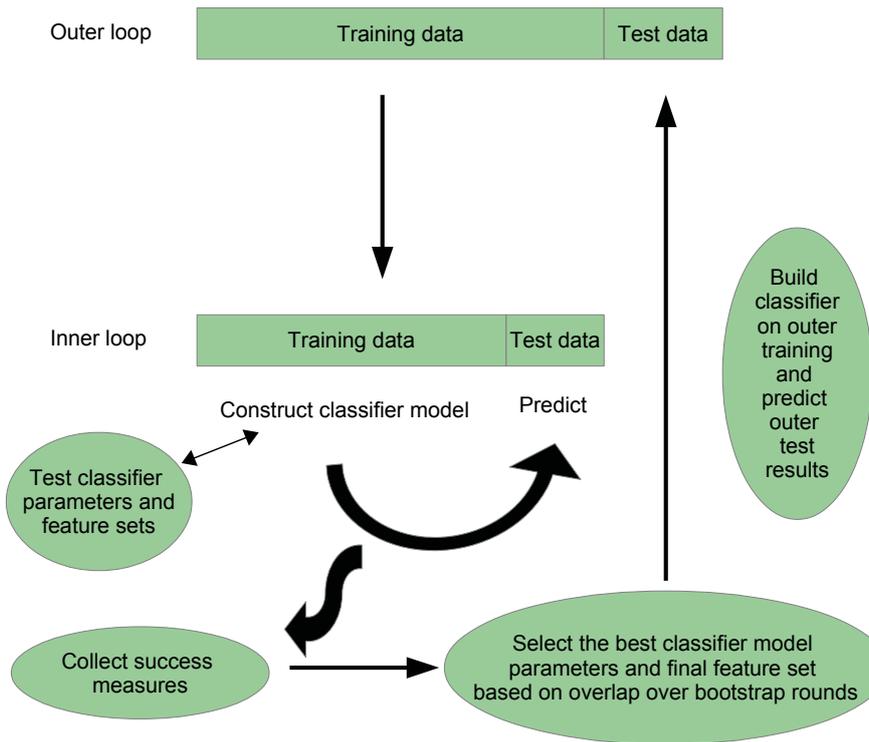


Figure 6.3: Nested cross validation: In the inner loop, the performance of different sets of classifier parameters and features are estimated based on a bootstrap cross-validation. The optimal classifier parameters and features are selected based on the performance evaluation over several bootstrap rounds. In the outer loop, model performance of the optimized classifier parameters and feature subset are evaluated on the hold-out test set. The outer loop is repeated several times, every time with potentially different classifier parameters and features.

132, 133].

A representation of classification performance that takes into account the successful as well as the unsuccessful classification results for both the positive and the negative class in a two class problem is the confusion matrix. The true positives (TP) are the instances that are correctly classified as belonging to the positive class and the false positives (FP) are the instances that are classified as the positive class but really belong to the negative class. The true negatives (TN) and false negatives (FN) can be explained similarly, see Table 6.1.

	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

Table 6.1: A confusion matrix with TP meaning true positive, FP meaning false positive, TN meaning true negative, and FN meaning false negative.

In many situations it will be an advantage to report a single classification performance measure instead of the whole confusion matrix which does not share the same sensitivity to the data imbalance problem as *accuracy* and *error rate*. Two common measures used in the machine learning community is precision and recall. Precision is a measure of exactness and is defined as the fraction of instances that are correctly classified to all instances that are classified as this class and is also known as positive predictive value. Recall is a measure of completeness and is defined as the fraction of instances that are correctly classified to all the instances that really belong to this class and is also known as true positive rate or sensitivity. Based on the confusion matrix shown above (Table 6.1), precision is then defined as: $Precision = \frac{TP}{TP+FP}$ and recall is defined as: $Recall = \frac{TP}{TP+FN}$.

Area under curve (AUC) is yet another performance measure used in the machine learning community. It is the calculated area under the receiver operating characteristics (ROC) graph [134, 135]. A ROC graph, is formed by plotting true positive rate (TP_{rate}) over false positive rate (FP_{rate}) which are defined as: $TP_{rate} = \frac{TP}{TP+FN}$ and $FP_{rate} = \frac{FP}{FP+TN}$. A classifier produces a TP_{rate} and an FP_{rate} which serve as a point in ROC space. By varying the decision threshold in a probabilistic classifier additional points in ROC space can be made making it possible to draw a graph between points. By doing the necessary steps, it is possible to draw ROC graphs from the results of hard-type classifiers as well [136, 137]. The ROC curve serves as a visual representation of trade-offs between benefits (TP) and costs (FP). Theoretically, the point $(0, 1)$ in ROC space is the result of the optimal classifier and the closer the ROC graph is to this point, the better the performance of your system. A common way of representing the results from ROC analysis as a single scalar value, is by calculating the area under curve (AUC). The closer the area is to 1.0 the better the performance of your system.

6.3 The imbalanced data problem

Studying data from a cohort, potentially introduce data groups with unequal number of instances as well as groups containing few instances. Most learning algorithms will fail to perform accurately under circumstances where data is imbalanced since balanced class distributions or equal misclassification costs are expected. The distributive characteristics of the data is often difficult to describe resulting in poor accuracy across data groups. In many cases the majority class will achieve close to 100 percent accuracy showing only 0-10 percent accuracy for the minority class. In a medical context where most of the instances are represented in the majority class as healthy subjects and only some of the instances are represented in the minority class as patients, this could have severe consequences since most patients would be wrongly classified as healthy subjects. The challenge is further increased when the dataset is small. The induction rules describing the minority group becomes fewer and weaker compared to the majority group. If, in addition, the size of the feature vector is large, learning algorithm easily gets too specific leading to overfitting and low generalization abilities.

A requirement in classification when the dataset is imbalanced is high accuracy for the minority class without simultaneously reducing the accuracy for the majority class too much. Many approaches have been suggested in the literature in an attempt to remedy the problem, see [138] for a thorough review. In this work we have focused on two different approaches, namely synthetic minority oversampling technique (SMOTE) [139] and cost-sensitive classification.

6.3.1 SMOTE

Even though it has been shown that classifiers learned from imbalanced data can perform well [140, 141], many classifiers performs better when trained on balanced data [142, 143, 144]. Sampling methods, such as over- and undersampling, have the potential to increase classification performance by either randomly undersample the majority class, oversample the minority class, or both. In the first case a drawback can be that important class concepts is missed during learning, in the second case replicated samples of certain instances can lead to overfitting and increases computation time. Combining both procedures can reduce some of the mentioned limitations.

The SMOTE algorithm creates synthetic data using information from feature space. 1) start with the feature vector from an existing instance x_i in the minority class, 2) randomly choose one of its K nearest neighbors \hat{x}_i , 3) calculate the difference vector between the instance under consideration and the neighbor $\hat{x}_i - x_i$, 4) multiply the difference vector by a random number δ between 0 and 1, and 5) add it to the vector under consideration $x_{new} = x_i + (\hat{x}_i - x_i) \times \delta$. The process is repeated for other x_i 's until the number of instances in the minority class has reached a satisfying level.

Creating synthetic data using SMOTE can be beneficial since the synthetic

Table 6.2: A cost matrix showing a numerical value for the cost related to the classification results. TP and TN are the number of correct classified positive and negative instances respectively and C_{TP} and C_{TN} is the cost related to correct classification. FP is the number of instances that were predicted to be positive, but were truly negative. FN is the number of instances that were predicted to be negative, but were truly positive. If the majority class is considered the positive class and the minority class is considered the negative class, then C_{FP} is typically set to 1 and C_{FN} to an arbitrary number n higher than one.

Cost matrix	True positive	True negative	
Predicted positive	C_{TP}	C_{FP}	=
Predicted negative	C_{FN}	C_{TN}	

0	1
n	0

instances avoids the ties introduced by replication to the learned classifier. On the other hand, a limitation is how SMOTE generates the same amount of new instances from each original sample which can increase overlapping between classes [145].

6.3.2 Cost-sensitive classification

As mentioned above, often in a medical context misclassification of the minority class has greater and more negative consequences compared to misclassification of the majority class. This is especially important to consider when the majority of instances in the dataset are healthy subjects and the minority of instances are patients. As an alternative to sampling, another approach to address these differences is by including an increased cost to misclassification of the minority class than the opposite into the learning algorithm.

In [146], McCarthy et al. discusses the performance of cost-sensitive learning and sampling techniques. They conclude that in most small datasets, sampling techniques show better performance and that in most, but not all, large datasets the situation is reversed.

The cost matrix is an essential concept in cost-sensitive learning. It is a numerical representation of misclassification penalties. In the upper left and lower right corners are the costs related to the true positive and true negative classification outcomes respectively. Mostly, there are no cost associated with correct classification, so C_{TP} and C_{TN} are typically zero. Since the cost related to misclassification of instances in the minority class are greater than the opposite, the C_{FP} is typically set to 1 and C_{FN} to an arbitrary number n higher than one, see Table 6.2. One solution to adding cost sensitivity to a probabilistic classifier, is by deciding on the class with lowest cost instead of the class with highest probability of correct classification. This is the same as minimizing the expected risk. In [147], Lomax reviews over 50 algorithms where the objective was to add cost sensitivity to classification trees.

Chapter 7

Contributions

After reading through chapters 1-6 the reader should have an understanding of the motivation behind this thesis and the objectives as well as the theoretic background for interpreting the results published in- and submitted to journals and published in conference proceedings. In the following chapter a summary of these results will be displayed organized per paper or manuscript.

7.1 Paper I

White Matter Hyperintensities in Mild Lewy Body Dementia, K. Oppedal, D. Aarsland, M.J. Firbank, H. Sønnesyn, O.B. Tysnes, J.T. O'Brien, M.K. Beyer, *Dement Geriatr Cogn Disord Extra*, 2012.

In Paper I, we studied the statistical differences concerning WML load in the periventricular regions as well as the deep white matter areas in MRI of the brain between two different dementia groups, an Alzheimer's disease group, a Lewy body dementia group, as well as a normal control group. MRI scans for 77 (61 AD and 16 LBD) patients and 37 healthy elderly control subjects were available for analysis. Baseline characteristics of the subjects can be found in Section 3.4 page 15 in Chapter 3. Nonparametric tests were used to compare groups. Additionally, the correlation between WML volume and cognition were calculated to explore the possible association between increased WML volume and cognition. Finally, we carried out multivariate regression to further explore this association and at the same time eliminate confounders.

7.1.1 Results for dementia vs. NC

The dementia group had numerically higher WML volumes than NC for total- and periventricular WML, but the differences did not reach statistical significance as shown in Table 7.1, see also Figure 7.1. Since the dementia groups were

Variable AD+LBD	Sub-group (dementia) n=37	Matched NC n=37	p-value dementia vs. nc
tot_WML%, median(iqr)	1.18 (1.88)	1.11 (0.87)	0.492 ^{MW}
pvh_WML%, median(iqr)	0.88 (1.29)	0.61 (0.61)	0.185 ^{MW}
dwm_WML%, median(iqr)	0.18 (1.29)	0.35 (0.68)	0.499 ^{MW}

Table 7.1: WML volumes (percent of brain volume) between a sub-group of demented (AD and LBD) matched for age, sex, and years of education to a group of normal controls. tot_WML% = total White Matter Lesion volume (percent of total brain volume), pvh_WML% = periventricular White Matter Lesion volume (percent of total brain volume), dwm_WML% = deep white matter White Matter Lesion volume (percent of total brain volume), iqr = inter quartile range, and MW = Mann Whitney U test.

not matched for sex, we examined differences in WML in the male subjects (14 AD and 13 LBD) separately - there were no differences between the male AD and LBD patients in either total- ($p=0.141$), periventricular- ($p=0.325$), or deep white matter WML ($p=0.202$).

7.1.2 Results for AD vs. LBD

Total-, periventricular-, and deep white matter WML volumes as percentage of total brain volume in the AD- and LBD groups are shown in Table 7.2 and Figure 7.2. There were no significant differences in WML volumes between the AD- and LBD groups (total WML: $p=0.238$, periventricular WML: $p=0.264$, and deep white matter WML: $p=0.444$), although LBD patients had higher numerical values for all three WML measures.

7.1.3 Results for correlation between WML volume and cognition

In the 77 dementia cases, there were significant correlations between cognitive scores and WML, see Table 7.3. We then analyzed AD and DLB groups separately. In the AD group, total-, periventricular-, frontal-, but not deep white matter WML correlated (Spearman rho, p-value) significantly with MMSE and verbal fluency, but not CVLT-2 (MMSE; total WML: $-(0.361, 0.004)$, periventricular WML: $(-0.296, 0.02)$, frontal WML: $(-0.392, 0.002)$, and fluency; total WML: $(-0.318, 0.013)$, periventricular WML: $(-0.278, 0.031)$, frontal WML: $(-0.376, 0.003)$). In contrast, the correlations were low and insignificant in the LBD group. In the AD group, there were correlations between years of edu-

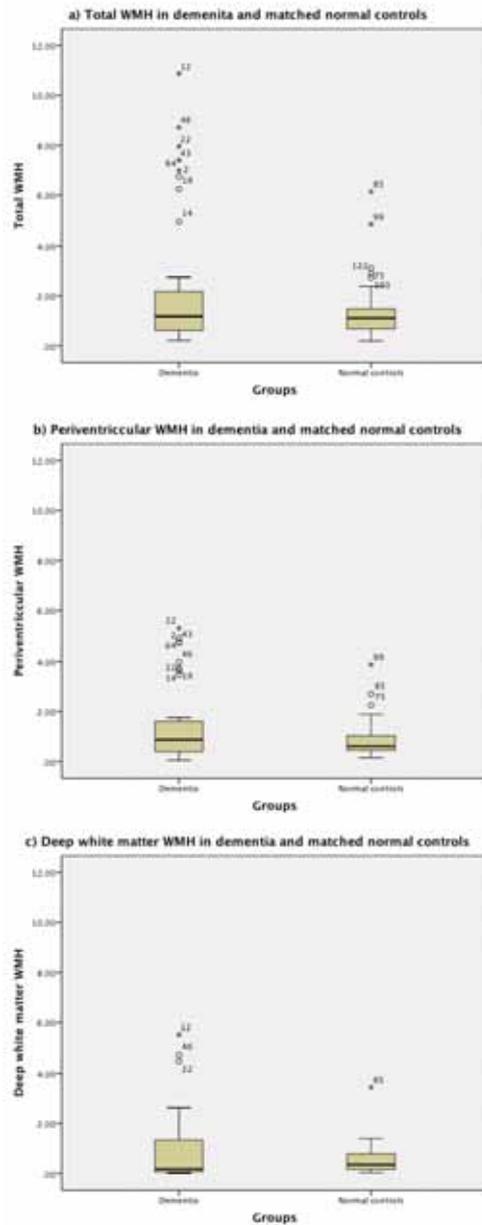


Figure 7.1: Box plots showing the distribution of WML volumes comparing the dementia group vs. normal controls (NC). a) shows total WML in dementia vs. NC, b) shows periventricular WML in dementia vs. NC, and c) shows deep white matter WML in dementia vs. NC.

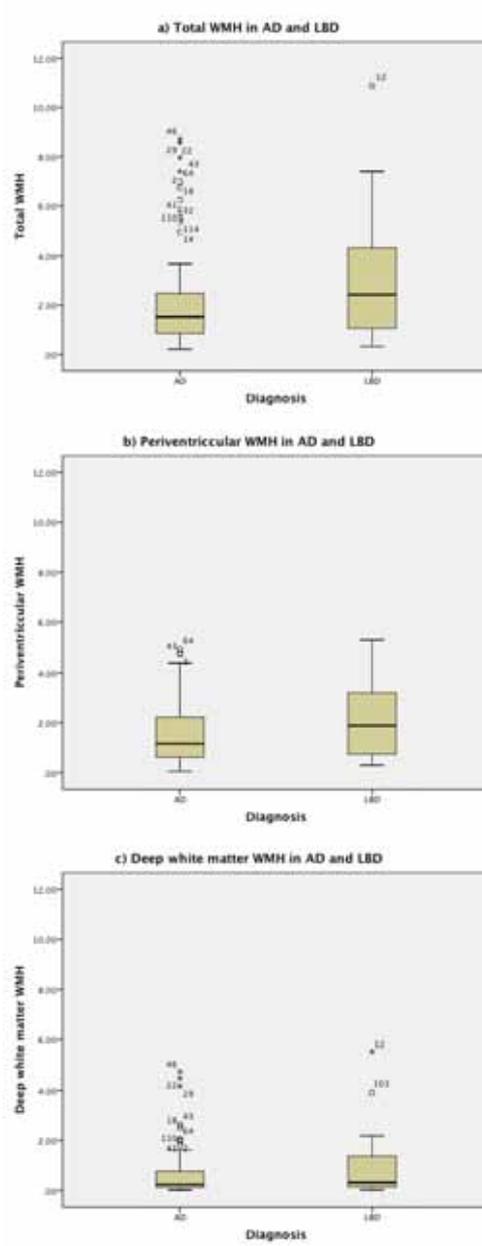


Figure 7.2: Box plots showing the distribution of WML volumes comparing AD and LBD. a) shows total WML in AD vs. LBD, b) shows periventricular WML in AD vs. LBD, and c) shows deep white matter WML in AD vs. LBD.

Variable	AD n=61	LBD n=16	p-value AD vs. LBD
tot.WML%, median(iqr)	1.53 (1.68)	2.43 (3.79)	0.238 ^{MW}
pvh.WML%, median(iqr)	1.16 (1.65)	1.89 (2.59)	0.264 ^{MW}
dwm.WML%, median(iqr)	0.24 (0.74)	0.31 (1.26)	0.444 ^{MW}

Table 7.2: WML volumes (percent of brain volume) between groups AD and LBD. tot.WML% = total White Matter Lesion volume (percent of total brain volume), pvh.WML% = periventricular White Matter Lesion volume (percent of total brain volume), dwm.WML% = deep white matter White Matter Lesion volume (percent of total brain volume), iqr = inter quartile range, and MW = Mann Whitney U test.

cation and MMSE, CVLT-2, and fluency. There were significant correlations between age and education and CVLT in the AD group, but not in the DLB group, see Table 7.3 for further details.

7.1.4 Results for multivariate regression

Multivariate regression was also applied to eliminate confounders, see Table 7.4. Hierarchical multiple regression was used to further study the association between WML (first total- and then frontal WML volume) and cognition (first MMSE and then verbal fluency). Total WML explained an additional 5.4% of the variance in MMSE, after controlling years of education. (R² change = 0.054, F change (1, 58) = 4.445, p = 0.039. Beta for total WML was -0.237, p = 0.039). Frontal WML explained an additional 4.9% of the variance in MMSE. R²change = 0.049, F change (1,58) = 4.030, p = 0.049. Beta for frontal WML was -0.230, p = 0.049). There were no significant contribution from WML on verbal fluency beyond that of years of education. See Table 5 for details.

7.1.5 Discussion

Using a semi-automatic method to measure volumes of WML on FLAIR images, similar levels of WML were found in patients with mild AD and LBD. WML load was associated with cognition in AD, but not in LBD, suggesting that although LBD have WML similar to AD, WML contributes to cognitive decline in AD, but not LBD.

Earlier studies have shown that WML volume is increased in AD patients compared to elderly people without dementia [25], but few previous studies of WML in LBD exist. In a study of patients with AD, PDD, DLB, and normal

Variable	AD	LBD
MMSE	n = 61	n = 16
tot_WML	-0.361 (0.004)	0.008 (0.978)
pvh_WML	-0.296 (0.020)	-0.048 (0.859)
dwm_WML	-0.227 (0.078)	0.183 (0.497)
frontal WML	-0.392 (0.002)	0.023 (0.933)
Age	-0.150 (0.249)	-0.056 (0.836)
Education	0.461 (0.000)	0.195 (0.468)
Sex	-0.080 (0.542)	-0.107 (0.692)
CVLT-2	n = 61	n = 14
tot_WML	-0.098 (0.451)	-0.002 (0.994)
pvh_WML	-0.171 (0.188)	-0.122 (0.679)
dwm_WML	0.078 (0.552)	0.164 (0.576)
frontal WML	-0.148 (0.254)	-0.064 (0.828)
Age	-0.316 (0.013)	-0.512 (0.061)
Education	0.324 (0.011)	-0.077 (0.794)
Sex	0.175 (0.177)	-0.239 (0.410)
Verbal Fluency	n = 60	n = 16
tot_WML	-0.318 (0.013)	0.101 (0.711)
pvh_WML	-0.278 (0.031)	0.145 (0.592)
dwm_WML	-0.210 (0.107)	0.220 (0.412)
frontal WML	-0.376 (0.003)	0.086 (0.752)
Age	-0.195 (0.136)	-0.211 (0.433)
Education	0.422 (0.001)	-0.101 (0.709)
Sex	-0.167 (0.203)	0.000 (1.000)

Table 7.3: Nonparametric bivariate correlations between WML volume and cognition score in the different groups of the study and different scores for cognition and between age and different scores for cognition. Numbers represent Spearman rho and p-values. tot_WML = total WML correlation coefficient, pvh_WML = periventricular WML correlation coefficient, and dwm_WML = deep white matter WML correlation coefficient.

Independent vs. dependent (controlling for education)	R Square	R Square Change	F Change (df1, df2)	Sig F. Change	Beta (Educ. WML)	Sig. (Educ. WML)
MMSE vs. tot WML	0.299	0.054	4.445 1.580	0.039	0.447 -0.237	0.000 0.039
MMSE vs. frontal WML	0.294	0.049	4.030 1.580	0.049	0.434 -0.230	0.000 0.049
fluency vs. tot WML	0.153	0.003	0.182 1.570	0.671	0.377 -0.053	0.004 0.671
fluency vs. frontal WML	0.158	0.008	0.535 1.570	0.467	0.363 -0.092	0.006 0.467

Table 7.4: Linear hierarchical regression with cognition (MMSE and fluency) as dependent, WML volume as independents, and controlling for years of education. tot WML = total White Matter Lesion volume, pvh WML = periventricular White Matter Lesion volume, and dwm WML = deep white matter White Matter Lesion volume.

controls, the authors did not find any differences between total WML, periventricular WML, or total deep WML in subjects with PDD or LBD compared to controls [148]. Subjects with AD had significantly greater volumes on all three measures compared to controls. This is in line with our results, although in our study, the difference in WML between dementia and NC groups did not reach significance, possibly related to the small sample size. Another possible explanation is the mild disease stage, since in a 1-year longitudinal follow-up, a significant increase in total WML within all the individual groups with the exception of the LBD group was found [148]. In a diffusion tensor imaging (DTI) study where fractional anisotropy (FA) values were significantly reduced in a number of white matter areas in the LBD group, no correlations between FA values and MMSE scores were found in the LBD subjects [149]. This could imply a different impact of WML in AD than LBD, consistent with our results.

There are methodological limitations that may have influenced our findings. Due to insufficient quality of scans, only a subset of available scans could be included, and thus the sample sizes were relatively small, in particular the LBD group. In addition LBD is a heterogeneous group. Thus, statistical power to detect minor differences may be reduced. In addition, the patients who were included had lower HIS and lower CDR-SOB scores, suggesting that they were slightly healthier than those excluded. Furthermore, this was a cross-sectional study, and thus we cannot conclude whether WML can cause cognitive decline.

The diagnosis was clinical, and misdiagnosis of AD and LBD cannot be excluded. However, standardized and validated clinical instruments were used, and patients were followed longitudinally. Among 20 patients with a clinical DLB diagnosis, 17 had abnormal uptake in the striatum on DaTscan. A small group

with autopsy diagnosis from this cohort ($n=7$), showed full consistence between pathological and clinical diagnosis. Thus, we believe the dementia diagnoses are accurate.

Several methods for quantification of WML severity from MR images of the brain exist [150, 151, 152]. A significant limitation to all of the above referenced methods is that they are prone to inter rater variability. It has been shown that quantification of the actual volume of WML is a more sensitive approach than the visual rating scales [153]. Others have found that visual rating is as good as the more complex methods in routine clinical practice, but that volumetric assessment should be used in research settings if possible [154].

Automated quantitative segmentation methods are suitable for assessing impact of WML on cognitive function [155]. Thus we decided to use an automatic segmentation method for the volume calculations [156], which is a strength of this study. Since the segmentation method was not robust enough to handle the variation in MR image quality in the DemWest multicenter study, we needed to manually edit the segmentation results to ensure correct results. The reliability of this manual editing between two raters, was ascertained. A more robust and fully automatic segmentation method such as [157], [158], or [159] might have increased the accuracy of our results further.

A possible limitation to our study was that we chose to calculate total brain size as the sum of gray matter and white matter, and use that as a point of reference when calculating lesion volume ratios. As age increases and possible dementia evolves, a significant atrophy is expected in both gray- and white matter [160, 161]. Others, [162], have used total intracranial volume as a reference point since lower degree of change is normally found here.

In multicenter studies differences between scanners and differences between imaging protocols are known to impact the results of automatic segmentation methods [163, 164, 165]. In our study we found it difficult to choose a common threshold level to automatically segment the WML without manual editing, probably caused by slight differences in imaging protocols. This was solved by choosing a method where manual editing was applied, to ensure that the reported volumes represent only WML and not other tissue with similar signal intensity such as fat.

To conclude, we found that although severity of WML did not differ significantly in mildly demented patients with AD and LBD, WML were associated with cognitive decline in AD, but not LBD. More studies of the potential clinical impact of WML in patients with LBD are needed.

7.2 Paper II

Using local binary pattern to classify dementia in MRI, K. Oppedal, K. Engan, D. Aarsland, M. Beyer, O.B. Tysnes, T. Eftestøl, *Proceedings of International Symposium on Biomedical Imaging (ISBI)*, 2012.

In the former paper (Paper I), it was shown that nonparametrical statistical tests couldn't reveal strong differences between WML volume in the three groups AD, LBD, and NC. A relevant question would be if further developing the analysis by applying a machine learning framework to WML volume features would be more potent regarding revealing differences between groups. In this paper, WML volume in different regions of the brain as well as texture analysis parameters were used as features in a machine learning framework with the aim to differentiate between a dementia group and normal controls. The total- and regional WML volumes in various regions were used as features. Total brain-, total periventricular-, and total deep white matter WML volume, as well as the temporal- (Te), occipital- (Oc), frontal- (Fr), parietal- (Pa), cerebellar- (CB), basal ganglia- (BG), central white matter (CW), and corpus callosum (CC) in each of the three regions were used to form 27 features. Mean, standard deviation, and entropy of the gray-scale intensities in the concatenation of all the WML areas of a patient, were used to form three other features. A maximum likelihood (ML) classifier was applied, see Section 6.2.1 page 39 in Chapter 6. Because of the small dataset, a leave-one-out technique was used as a strategy for cross-validation, see Section 6.2.3 page 44 in Chapter 6. Area under receiver operating curve (AUC), was used to quantify the performance of the classifier, see Section 6.2.4 page 45 in Chapter 6. All texture features were calculated from the FLAIR MR image.

7.2.1 Results for 1D feature vectors

In the first experiment, all possible 1D feature vectors were tested classifying the demented from the normal controls. The best classification result using the WML volumes as features were total cerebellar WML volume with an AUC of 0.73 and 95% confidence interval of 0.57 to 0.83. Similar results were achieved for the cerebellar WML volume in the deep white matter, the frontal WML volume in the periventricular region and the volume of WML in the basal ganglia, see Table 7.5.

The AUC from the mean, standard deviation, and entropy of the grayscale values in the ROI were all close to or below 0.6. Because of the unpromising result, these features were not further used. Using the same classification framework texture analysis features were applied. The decision fell on a texture feature named local binary pattern (LBP) see Section 5.3 page 29 in Chapter 5. The highest AUC values from the LBP features, were median value of the all the LBP values calculated with radius two using eight samples and similar with

Feature	AUC (CI_{95})
CB_{tot}	0.73 (0.57, 0.83)
CB_{dwm}	0.73 (0.57, 0.83)
Fr_{pvh}	0.73 (0.54, 0.86)
BG_{dwm}	0.71 (0.52, 0.84)
BG_{pvh}	0.71 (0.51, 0.84)
pvh	0.70 (0.51, 0.84)

Table 7.5: The six 1D volume feature vectors giving the highest AUC classifying normal controls and demented.

radius four using 16 samples with AUC equal to 0.86 (0.68, 0.93) and 0.82 (0.64, 0.91) respectively, see Table 7.6.

7.2.2 Results for 2D feature vectors

The best performing 1D feature vectors were combined with all other possible 1D feature vectors to generate 2D feature vectors which were tested in the same manner as above, and provided a significant improvement over 1D feature vectors. The 2D feature vector with the highest AUC were the combination of the entropy of all the calculated LBP values with radius four and 16 samples combined with the median of all the calculated contrast measures with radius eight and 16 samples with AUC equal to 0.90 (0.74, 0.96), see Table 7.7.

7.2.3 Results for 3D feature vectors

In the same manner 3D feature vectors were made combining the best performing 2D feature vector with all other 1D feature vectors only performing marginally better than the best 2D feature vectors. The 3D feature vector with the highest AUC were the combination of the entropy of all the calculated LBP values with radius four and 16 samples, median of all the calculated contrast measures with radius eight and 16 samples, and the median of all the calculated LBP values with radius eight and 16 samples with AUC equal to 0.91 (0.75, 0.97), see Table 7.8.

7.2.4 Discussion

Based on image processing of WML in MR images of demented and normal controls, the LBP features used in our experiment are proposed as powerful features in a maximum likelihood classifier, when classifying demented from normal controls.

Earlier Kloppel et al. [166] successfully used whole brain images as features in a SVM classifier to distinguish subjects suffering from Alzheimer disease from healthy elderly. To our knowledge though, regional WML volumes, LBP, and C

Feature	AUC (CI_{95})
LBP_{median}^{r2p8}	0.86 (0.68, 0.93)
LBP_{median}^{r4p16}	0.82 (0.64, 0.91)
LBP_{mean}^{r1p8}	0.79 (0.61, 0.90)
$LBP_{entropy}^{r4p16}$	0.79 (0.61, 0.90)
$LBP_{entropy}^{r2p8}$	0.79 (0.60, 0.90)
LBP_{mean}^{r2p8}	0.77 (0.58, 0.89)
$LBP_{skewness}^{r2p8}$	0.75 (0.56, 0.88)
$LBP_{skewness}^{r1p8}$	0.75 (0.56, 0.88)
C_{mean}^{r8p16}	0.75 (0.56, 0.87)
$LBP_{entropy}^{r1p8}$	0.74 (0.55, 0.86)
C_{median}^{r8p16}	0.72 (0.53, 0.85)
$LBP_{entropy}^{r8p16}$	0.70 (0.52, 0.83)

Table 7.6: The twelve 1D texture feature vectors giving the highest AUC classifying normal controls and demented.

Features	AUC (CI_{95})
$LBP_{entropy}^{r4p16}, C_{median}^{r8p16}$	0.90 (0.74, 0.96)
$LBP_{mean}^{r2p8}, C_{median}^{r8p16}$	0.89 (0.72, 0.96)
$LBP_{median}^{r2p8}, C_{mean}^{r8p16}$	0.89 (0.72, 0.96)
$LBP_{mean}^{r1p8}, C_{median}^{r8p16}$	0.89 (0.72, 0.96)
$LBP_{mean}^{r1p8}, C_{mean}^{r4p16}$	0.88 (0.71, 0.95)
$LBP_{entropy}^{r4p16}, C_{mean}^{r8p16}$	0.88 (0.71, 0.96)
$LBP_{entropy}^{r8p16}, C_{median}^{r8p16}$	0.88 (0.71, 0.96)
$LBP_{median}^{r2p8}, C_{mean}^{r4p16}$	0.88 (0.71, 0.95)

Table 7.7: The eight 2D texture feature vectors giving the highest AUC classifying normal controls and demented.

Feature	$AUC(CI_{95})$
$LBP_{entropy}^{r4p16}$, C_{median}^{r8p16} , LBP_{median}^{r8p16}	0.91 (0.75, 0.97)
$LBP_{entropy}^{r4p16}$, C_{median}^{r8p16} , $LBP_{kurtosis}^{r8p16}$	0.90 (0.74, 0.96)

Table 7.8: The two 3D texture feature vectors giving the highest AUC classifying normal controls and demented.

values calculated from WML regions in MR images have not earlier been used as features to classify demented from normal controls. Since the distribution of the features in the dataset may not be normal, this approach can underestimate the true AUC values.

The small dataset of 95 subjects is a limitation of the study, but a leave-one-out cross-validation technique is adopted to exploit the dataset as good as possible. It is not certain that the features we used are normally distributed, and thus it is possible that the results would improve if a nonparametric classification technique was adopted.

7.3 Paper III

Classifying dementia using local binary patterns from different regions in magnetic resonance images, K. Oppedal, T. Eftestøl, K. Engan, M. Beyer, D. Aarsland, *International Journal of Biomedical Imaging*, 2015.

Based on the promising results from the texture analysis approach in Paper II, further investigation was deemed required. Relevant questions would be if texture features calculated from the T1 images could improve classification of dementia further, especially the LBD class, compared to the FLAIR images. And how will increasing the size of the region of interest (ROI) into all white matter influence performance?

We wanted to test if the WML regions inherit textural information in an extent that can be used to classify dementia patients from normal controls and AD from LBD. As the detection of textural information in WML might not be dependent on an exact delineation of WML, we also wanted to test if a comparable classification accuracy can be achieved using all of WM as ROI, since WM segmentation is more available and only a 3DT1 MR image is needed which is commonly acquired in a clinical setting. We wanted to test different types of LBP calculations together with a contrast measure (C) calculated from FLAIR and 3DT1 MR images and on a subset containing data from one scanner only.

Because of the challenging situation with imbalanced data having different numbers of subjects in the represented groups in the above mentioned cohorts, we wanted to test how the use of resampling of instances affect classification results. For that purpose SMOTE, see Section 6.3.1 page 48 in Chapter 6, was applied to the data.

We decided to apply the random forest classifier since it is known to be robust showing high classification performance with low bias and variance in many classification problems, see Section 6.2.2 page 41 for more information.

7.3.1 Results for three class problem, NC vs. AD vs. LBD

Results for the three-class problem with class 0 being NC, class 1 being AD and class 2 being LBD are shown in detail in Table 7.9. TotAcc is the total accuracy for all three classes. P0 is the precision for the NC group, P1 is the precision for the AD group, P2 is the precision for the LBD group, R0 is the recall for the NC group, R1 is the recall for the AD group, and R2 is the recall for the LBD group.

The first test named FLAIR-WML_{ri} indicate that the FLAIR MR image was used for calculation of LBP and C, that WML was the ROI, and that the rotational invariant variant of the LBP feature was used. The second test named T1WML_{ri} indicates that the T1 MR images was used for calculation of LBP and C, that WML was the ROI, and that the rotational invariant variant of the LBP

Test	TotAcc	P0 R0	P1 R1	P2 R2
FLAIR-WML _{ri}	0.60(0.13)	0.71(0.28) 0.48(0.25)	0.61(0.14) 0.77(0.28)	0.33(0.41) 0.20(0.35)
T1WML _{ri}	0.82(0.12)	0.96(0.10) 0.98(0.08)	0.80(0.11) 0.88(0.18)	0.58(0.49) 0.25(0.35)
T1WML_{ri}^{SMOTE}	0.87(0.08)	0.97(0.07) 1.00(0.00)	0.81(0.17) 0.82(0.16)	0.85(0.11) 0.78(0.20)
T1WM _{ri}	0.82(0.09)	0.96(0.08) 1.00(0.00)	0.81(0.11) 0.88(0.16)	0.42(0.49) 0.20(0.35)
T1WM _{ri} ^{SMOTE}	0.75(0.13)	0.90(0.12) 1.00(0.00)	0.66(0.16) 0.72(0.19)	0.70(0.21) 0.55(0.22)
T1WML _{svg,ri} ^{SMOTE}	0.91(0.15)	1.00(0.00) 1.00(0.00)	1.00(0.00) 0.77(0.42)	0.87(0.22) 1.00(0.00)

Table 7.9: Results are reported as mean with standard deviation in brackets, m(s), over 10 folds cross validation, classifying NC vs. AD vs. LBD. TotAcc=total accuracy, R=recall, P=precision. 0 for class NC, 1 for class AD, and 2 for class LBD. ROI is either WM for white matter or WML for white matter lesion area.

feature was used. The third test named T1WM_{ri} indicates that the T1 MR images was used for calculation of the LBP and C, that the WM was the ROI, and that the rotational invariant variant of the LBP feature was used.

The total accuracy showed great variation throughout the different tests ranging from 0.6(0.13) to 0.87(0.08). The performance increased considerably when calculating the LBP and C features from the T1 MR image as compared to the FLAIR MR image. The classification performance proved best in the T1 case and when WML was used as ROI.

For comments on the T1WML_{svg,ri}-test, see Section 7.3.4.

7.3.2 Results for the two class problem, NC vs. AD+LBD

Results for the two-class problem with class 0 being NC and class 1 being AD and LBD together are shown in detail in Table 7.10. TotAcc is the total accuracy for the two classes. P0 is the precision for the NC group and P1 is the precision for the combined AD and LBD group, R0 is the recall for the NC group and R1 is the recall for the combined AD and LBD group.

In addition to the above mentioned tests, another test named T1WML_{riu2} was applied to assess whether the classification performance would differ when rotational invariant LBP were calculated alone or in combination with selection of uniform LBP values only.

Total accuracy is generally higher in the T1 case (ranging from 0.97(0.04) to

Test	TotAcc	P0 R0	P1 R1
FLAIR-WML _{ri}	0.80(0.12)	0.69(0.20) 0.72(0.23)	0.87(0.11) 0.84(0.12)
T1WML_{ri}	0.98(0.04)	0.98(0.06) 0.98(0.08)	0.99(0.04) 0.99(0.05)
T1WM _{ri}	0.97(0.04)	0.96(0.08) 0.98(0.08)	0.99(0.04) 0.97(0.06)
T1WML _{riu2}	0.98(0.04)	0.96(0.08) 1.00(0.00)	1.00(0.00) 0.97(0.06)
T1WML _{svg,ri}	1.00(0.00)	1.00(0.00) 1.00(0.00)	1.00(0.00) 1.00(0.00)

Table 7.10: Results are reported as mean with standard deviation in brackets, m(s), over 10 folds cross validation, classifying NC vs. AD+LBD. TotAcc=total accuracy, R=recall, P=precision. 0 for class NC and 1 for class AD+LBD. ROI is either WM for white matter or WML for white matter lesion area.

0.98(0.04)) compared to the FLAIR case (0.80(0.12)), but approximately similar for the two different ROI's when T1 MR images are used. Precision for class 0 is higher in the case of LBP and C calculated in the WML area of the T1 image (0.98(0.06)) as compared to all of the WM area (0.96(0.08)). Recall for class 0 is similar for both ROI's. This is also the case for precision for class 1 (0.99(0.04)), but recall for class 1 is higher when LBP and C are calculated in the WML region 0.99(0.05) as compared to the WM region (0.97(0.06)).

When the rotational invariant calculation of LBP is combined with selection of the uniform values only, the P0 and R1 are similar to the *ri*-case. The *riu2*-case had marginally higher values for total accuracy, P1, and R0.

For comments on the T1WML_{svg,ri}-test, see Section 7.3.4.

7.3.3 Results for the two class problem, AD vs. LBD

Results for the two class problem with class 1 being AD and class 2 being LBD are shown in detail in Table 7.11.

Classification performance was highest in the T1 case when WM was used as ROI.

7.3.4 Results when using data from Stavanger only

In both the three-class problem and in the two-class problem NC vs. AD+LBD a fifth test was run named T1WML_{svg,ri} which indicates that the T1 MR images was used for calculation of the LBP and C, that the WM was the ROI, and that only data from the MR-scanner located at Stavanger University Hospital were used. This experiment was done to ensure that the results we achieved, were not

Test	TotAcc	P1 R1	P2 R2
FLAIR-WML _{ri}	0.73(0.15)	0.78(0.11) 0.91(0.12)	0.20(0.45) 0.10(0.32)
T1WML _{ri}	0.66(0.17)	0.74(0.10) 0.84(0.18)	0.00(0.00) 0.00(0.00)
T1WML^{SMOTE}_{ri}	0.73(0.16)	0.72(0.18) 0.75(0.20)	0.76(0.17) 0.71(0.19)
T1WM _{ri}	0.74(0.16)	0.80(0.09) 0.75(0.20)	0.45(0.51) 0.71(0.19)
T1WM ^{SMOTE} _{ri}	0.68(0.14)	0.67(0.14) 0.69(0.29)	0.75(0.21) 0.68(0.14)

Table 7.11: Results are reported as mean with standard deviation in brackets, m(s), over 10 folds cross validation, classifying AD vs. LBD. TotAcc=total accuracy, R=recall, P=precision. 1 for class AD and 2 for class LBD. ROI is either WM for white matter or WML for white matter lesion area.

a profit of the fact that the data were collected from different MR centers using slightly different MR protocols. The rotational invariant LBP feature was used in this test. An even better performance was reached in both cases. In the three-class problem a total accuracy of 0.91(0.15) was achieved and all of the cases in the dataset were classified correctly in the two-class problem. An implication of this is that between-center noise falsely reduces classification accuracy and that the developed method shows even higher performance when all data come from the same scanner.

7.3.5 Discussion

Our results improved doing LBP texture analysis in 3DT1 image rather than the FLAIR image, indicating that there exists more textural information in the 3DT1 image compared to the FLAIR image relevant to our problem formulation. In the three-class problem as well as in the two-class problem NC vs. AD+LBD our results indicates that there exists similar amount of relevant textural information regarding dementia classification using all of WM as ROI compared to using only WML. This could be a benefit. WML segmentation is unsatisfactory developed and very often demanding manual outlining is required as well as a FLAIR MR image, where WML is hyper intense, while WM segmentation is readily available from many well known and freely downloadable software packages needing only a 3DT1 MR image which is a common part of a clinical MR protocol. In addition recent focus on diffusion tensor imaging (DTI) in vascular disease [167], amnesic mild cognitive impairment (aMCI) [168], and dementia [169, 170, 68] strengthens the view that age-related changes in WM plays an important role in the development of dementia. DTI is never the less not suffi-

ciently available and at the same time costly making other approaches for WM analysis, like ours, a valuable addition.

In the two-class problem AD vs. LBD, we did not reach a comparable classification result compared to the AD+LBD vs. NC case. There probably exist several explanations for that, one of the most obvious being the small sample size in the LBD class compared to the other classes. The LBD subjects are mainly classified as AD subjects indicating that the two groups experience similarities concerning our methods. Even though the two groups show different neurological etiologies they do not differ equally regarding vascular changes. Having few subjects in the LBD group, the calculated texture features may not represent the group with proper specificity or generality. Another explanation could be related to the common basis for neurodegenerative dementias pointed out by Bartzokis in [171] or Schneider’s observations about mixed brain pathologies in dementia [172].

In the three-class problem NC vs. AD vs. LBD, is the accuracy for the LBD class improved showing a precision of 0.85(0.11) and recall of 0.78(0.20). When doing the same test on the data from Stavanger only, even better results were achieved with a precision of 0.87(0.22) and a recall of 1.00(0.00) for the LBD class. Vemuri et al. [62] used atrophy maps and a k-means clustering approach to diagnose AD with a sensitivity of 90.7% and a specificity of 84%, LBD with a sensitivity of 78.6% and specificity of 98.8%, and FTLD with a sensitivity of 84.4% and a specificity of 93.8%. A strength of their study was that they only used MR images of later histological confirmed LBD patients. They also report sensitivity and specificity for the respective clinical diagnoses. AD with a sensitivity of 89.5% and a specificity of 82.1%, LBD with a sensitivity of 70.0% and specificity of 100.0% and, FTLD with a sensitivity of 83.0% and a specificity of 95.6%. Compared to the reported sensitivity and specificity for clinical diagnosis, our method shows substantial higher accuracy for LBD and comparable accuracy for AD. A limitation is the use of different measures of goodness to the classification results and that different data is used. In [106] Kodama and Kawase a classification accuracy of 70% for the LBD group from AD and NC is reported. Burton et al. report a sensitivity of 91% and a specificity of 94% using calculations of medial temporal lobe atrophy assessing diagnostic specificity of AD in a sample of patients with AD, LBD and, vascular cognitive impairment, but do not report results for the LBD group [61]. In [65], Lebedev et al. uses sparse partial least squares (SPLS) classification of cortical thickness measurements reporting a sensitivity of 94.4 and a specificity of 88.89 discerning AD from LBD.

To verify that the classification results are not driven by differences in the local variation of signal intensities (the C values) between centers used during collection of MR data in the study, the test $T1WML_{svg,ri}$ was conducted on the Stavanger data only. The results showed an increase in classification performance, which gives us reason to believe that the results reflect real diagnostic differences.

LBP is based on local gradients and is therefore prone to noise and could be a limitation to our approach. LBP values calculated in a noisy neighbourhood would be recognized by many transitions between 0's and 1's. We performed a test, the $T1WML_{riu2}$ test, where only rotational invariant and uniform LBP values, showing a maximum of two transitions between 0's and 1's, are collected. The result showed identical results as the $T1WML_{ri}$ -test indicating that noise do not constitute a severe problem in our method. Even though noise reduction procedures can be useful in the application of for example segmentation, a noise reduction approach could remove relevant textures. The contrast measure is invariant to shifts in gray-scale but not invariant to scaling. We do not use any normalization of the images prior to the feature calculation. Thus, one could argue that different patients are scaled differently making the contrast measure less trustworthy. On the other hand, if a normalization is done, for example based on a maximum intensity value, this could indeed change the local subtle textures, and effect the contrast measures, possibly in a negative way. In the present work we have investigated the discriminating power of the features calculated without any smoothing or normalization, since the effect of such operators are not clear for this application. In future work we want to investigate the use of different preprocessing steps, both using denoising and/or normalization and compare the discrimination power of the features with and without preprocessing. The improvement in results when using data from one center only (Stavanger), indicates lack of robustness which can be related to the facts mentioned above.

Cronbach's alpha was calculated using total brain volume to ensure that our data material was consistent even though it was collected from different centers spanning a time scale. Texture features can be exposed to noise and a limitation to our study is the lack of using texture features for the reliability analysis.

Another limitation to our study is the lack of clinical interpretation of texture features which is difficult in our case, since brain regional information is lost in the process of feature calculation.

This study demonstrates that LBP texture features combined with the contrast measure C calculated from brain MR images are potent features used in a machine learning context for computer based dementia diagnosis. The results discerning AD+LBD from NC is especially promising, potentially adding value to the clinical diagnose. In the three-class problem, the classification performance exceeded the accuracy of clinical diagnosis for the LBD group and at the same time keeping the classification accuracy for the AD group comparable to the clinical diagnose. A lower accuracy was achieved when classifying AD from LBD in the two-class problem AD vs. LBD. We considered it good news that the results using WM as ROI gave almost equally good classification performance as WML, since the WM segmentation routine is much more accessible compared to WML segmentation. The performance using 3DT1 images for texture analysis was notably better than when using FLAIR images, which is an advantage, since most common MR protocols include a 3DT1 image.

7.4 Paper IV

Classifying Alzheimer’s disease, Lewy body dementia, and normal controls using 3D texture analysis in magnetic resonance images, K. Oppedal, K. Engan, T. Eftestøl, M. Beyer, D. Aarsland, *Submitted manuscript*, 2015.

In the previous study we concluded that 2D texture analysis calculated from WM and WML regions in 3DT1 MR images of the brain used in a random forest classifier, is able to classify subjects with dementia from healthy controls with high accuracy. Lower performance was reached when classifying subject with AD from subjects with LBD. With an ambition to increase classification performance for the LBD group we decided to try out 3D texture analysis. The main objective of this paper was to test whether 3D LBP texture features used in a random forest classifier is able to discern patients with AD and patients with LBD from normal controls with similar or higher accuracy than 2D LBP. Two different approaches to 3D LBP analysis were considered, (1) volume LBP (VLBP) and (2) LBP three-orthogonal-planes (LBP-TOP). Four classification tasks were tested, a three-class problem NC vs. AD vs. LBD, and three two-class problems, NC vs. AD, NC vs. LBD, AD vs. LBD. A second objective was to consider brain regional importance of LBP features in dementia classification by calculating texture features in the white matter lesion (WML) and normal appearing white matter (NAWM) areas of the brain. The data used in this study is imbalanced concerning the number of subjects in each group. In addition to performing classification on the original data, two methods for handling the imbalance problem were tested. One is to use synthetic minority oversampling technique (SMOTE) to balance the groups evenly and another is to use cost sensitive classification by adding a cost to the wrongly classified subjects. We wanted to elaborate on the clinical relevance of 3D LBP texture features by calculating correlation between features and cognition measured by mini mental state examination (MMSE) controlling for age.

In the tables, a naming convention was used for the name of the tests as follows:

$$ClassProb_{ROI,texture}^{data,ntrXX}$$

where *ClassProb* can be either *NC/AD/LBD*, *NC/AD*, *NC/LBD*, *AD/LBD*, *ROI* can be either *WML* or *NAWM*, *texture* can be either *VLBP* or *TOP*, *data* is either *orig*, *smote*, or *cost*, where *orig* means that no resampling was used, *smote* means that the smaller classes are resampled up to the same number as the largest class using SMOTE, and *cost* means that cost-sensitive classification was used, *ntrXX* refers to the number of trees used in the random forest classifier. In the case of the three-class problem, *ClassProb* has been omitted from the test name to save space.

7.4.1 Results - classification of NC, AD, and LBD

In this experiment we were looking at the three-class problem NC vs. AD vs. LBD for different textural features (Step 1), regions of interest (Step 2), using both the original data, resampling with SMOTE, or cost-sensitive classification (Step 3), and using different number of trees in the random forest classifier (Step 4). The best results are reported in Table 7.12. The number of subjects in each class is as follows: NC = 36, AD = 58, and LBD = 16.

Test	TotAcc	P0 R0	P1 R1	P2 R2
<i>smote, ntr50</i> <i>VLBP, WML</i>	0.78(0.10)	0.95(0.09) 0.93(0.14)	0.72(0.16) 0.69(0.11)	0.72(0.11) 0.73(0.19)
<i>smote, ntr20</i> <i>VLBP, NAWM</i>	0.78(0.10)	0.86(0.14) 0.92(0.12)	0.77(0.10) 0.69(0.25)	0.75(0.18) 0.73(0.15)
<i>smote, ntr30</i> <i>TOP, WML</i>	0.76(0.09)	0.86(0.14) 0.92(0.09)	0.75(0.18) 0.61(0.24)	0.73(0.14) 0.78(0.11)
<i>smote, ntr50</i> <i>TOP, NAWM</i>	0.79(0.07)	0.91(0.10) 0.97(0.07)	0.71(0.12) 0.69(0.13)	0.79(0.16) 0.71(0.18)

Table 7.12: Results for the three-class problem NC vs. AD vs. LBD are reported as mean with standard deviation in brackets, over 10 folds cross validation, classifying NC vs. AD vs. LBD. TotAcc=total accuracy, R=recall, P=precision. 0 for class NC and 1 for class AD and 2 for class LBD. ROI is either *WML* for white matter lesion area or *NAWM* for normal appearing white matter area. *TOP* is short for LBP-TOP. *smote* means that the data used for analysis was resampled. *ntrXX* refers to the number of trees used in the random forest classifier.

The best result was achieved when using LBP-TOP/C in the NAWM area and 50 trees in the random forest classifier. There were no benefit in adding cost-sensitive classification doing this experiment. In all cases resampling the data using SMOTE increased the precision and recall for the LBD class considerably and on the expense of total accuracy, precision and recall for the AD class, and in most cases for the NC class as well.

7.4.2 Results - three two-class problems

In this experiment we report results from the three different two-class problems (i, ii, and iii) through steps 1-4 in the same manner as above. i) The NC (36 instances) vs. AD (58 instances) problem shows similar results for both texture types and ROI's. The groups are not dramatically unbalanced and the results using either the original data, resampling with SMOTE, or doing cost-sensitive classification show very similar classification accuracy. Thus in Table 7.13, the

best results are shown and are marked with test name *NC/AD*. ii) Test NC vs. LBD (16 instances) showed best results when using VLBP texture calculated in the WML area. This is a more imbalanced test and using SMOTE gave best results in all cases. See the entrances in Table 7.13 marked with test name *NC/LBD*. iii) The test AD vs. LBD gave best results when VLBP texture was calculated in the NAWM region. The groups in this test is dramatically imbalanced. Here the precision and recall for the LBD group becomes lower than chance using the original data and resampling is necessary. The results using SMOTE are shown in Table 7.13 and are marked with test name *AD/LBD*.

7.4.3 Results - robustness test

We wanted to test the robustness of the proposed method by performing the classification on a subset of the original data where all the subjects were scanned on one scanner only. This subset includes 12 NC, 47 AD and 11 LBD instances. In the case of poor robustness large variation between classification results based on the original data and data from one scanner only would be large. In the case of a high degree of robustness the results would not differ too much.

In the highly imbalanced three-class problem (see Table 7.14) as well as the two-class problem AD vs. LBD (see Table 7.15), the results using a single scanner only (marked *ss* in the tables) are very similar to what is achieved using the entire set (marked *orig* in the tables) for all tests, hence a reasonable robustness is achieved.

In the classification of the two-class problems NC vs. AD and NC vs. LBD, the test using data from a single scanner achieve a classification accuracy of 100% in all tests which in most cases is close to the results achieved using the entire dataset which indicates a reasonable robustness as well.

7.4.4 Results - correlation of features vs. cognition

For the NC group, one VLBP-feature showed significant correlation in the WML area, none of the C-features and neither VLBP nor C showed correlations in the NAWM area. In the LBP-TOP/C-case, 10 out of 72 LBP-TOP-features showed significant correlations and one out of 72 C-features.

In the correlation tests the AD group stands out when features were calculated in the NAWM area. In the VLBP/C case, seven out of eight VLBP-features and one out of eight C-features showed significant correlation and in the LBP-TOP/C case, 60 out of 72 LBP-TOP-features and 37 out of 72 C-features showed significant correlations. When using WML as area for feature calculation only one VLBP-feature and two C-features in the LBP-TOP/C case showed significant correlation.

None of the features showed correlation for the LBD group.

Test	TotAcc	P0 R0	P1 R1
$NC/AD^{cost0120, ntr30}_{VLBP, WML}$	0.94(0.09)	0.93(0.11) 0.93(0.17)	0.96(0.10) 0.95(0.09)
$NC/AD^{orig, ntr30}_{VLBP, NAWM}$	0.97(0.07)	0.95(0.11) 0.98(0.08)	0.98(0.05) 0.97(0.07)
$NC/AD^{orig, ntr10}_{TOP, WML}$	0.97(0.07)	0.96(0.10) 0.98(0.08)	0.98(0.05) 0.97(0.07)
$NC/AD^{smote, ntr10}_{TOP, NAWM}$	0.94(0.07)	0.93(0.10) 0.97(0.07)	0.97(0.07) 0.92(0.12)
		P0 R0	P2 R2
$NC/LBD^{smote, ntr20}_{VLBP, WML}$	0.97(0.06)	1.00(0.00) 0.94(0.12)	0.96(0.10) 1.00(0.00)
$NC/LBD^{smote, ntr20}_{VLBP, NAWM}$	0.87(0.11)	0.86(0.16) 0.92(0.14)	0.93(0.12) 0.85(0.17)
$NC/LBD^{smote, ntr20}_{TOP, WML}$	0.89(0.13)	0.92(0.15) 0.90(0.17)	0.90(0.16) 0.90(0.17)
$NC/LBD^{smote, ntr10}_{TOP, NAWM}$	0.93(0.09)	0.97(0.11) 0.92(0.14)	0.94(0.11) 0.95(0.16)
		P1 R1	P2 R2
$AD/LBD^{smote, ntr30}_{VLBP, WML}$	0.76(0.13)	0.79(0.17) 0.74(0.17)	0.76(0.13) 0.78(0.19)
$AD/LBD^{smote, ntr50}_{VLBP, NAWM}$	0.79(0.15)	0.83(0.17) 0.74(0.25)	0.79(0.17) 0.83(0.16)
$AD/LBD^{smote, ntr30}_{TOP, WML}$	0.77(0.10)	0.80(0.14) 0.74(0.12)	0.76(0.12) 0.79(0.15)
$AD/LBD^{smote, ntr50}_{TOP, NAWM}$	0.72(0.18)	0.73(0.18) 0.73(0.25)	0.73(0.20) 0.72(0.24)

Table 7.13: Results for the three two-class problem NC vs. AD, NC vs. LBD, and AD vs. LBD are reported as mean with standard deviation in brackets, over 10 folds cross validation. TotAcc=total accuracy, R=recall, P=precision. 0 for class NC and 1 for class AD and 2 for class LBD. ROI is either *WML* for white matter lesion area or *NAWM* for normal appearing white matter area. *TOP* is short for LBP-TOP. *orig* means that the original untouched data was used for texture feature calculation, *cost* that cost-sensitive classification was applied, and *smote* that the data was resampled. *ntrXX* refers to the number of trees used in the random forest classifier.

Test	TotAcc	P0 R0	P1 R1	P2 R2
<i>orig, ntr20</i> <i>VLBP, WML</i>	0.86(0.10)	0.96(0.08) 0.97(0.11)	0.84(0.13) 0.93(0.12)	0.79(0.39) 0.45(0.44)
<i>ss, ntr20</i> <i>VLBP, WML</i>	0.84(0.11)	1.00(0.00) 1.00(0.00)	0.86(0.11) 0.94(0.11)	0.50(0.50) 0.30(0.48)
<i>orig, ntr50</i> <i>VLBP, NAWM</i>	0.82(0.14)	0.89(0.19) 0.92(0.14)	0.80(0.14) 0.93(0.09)	0.75(0.50) 0.25(0.42)
<i>ss, ntr50</i> <i>VLBP, NAWM</i>	0.79(0.14)	1.00(0.00) 1.00(0.00)	0.82(0.11) 0.88(0.14)	0.29(0.49) 0.20(0.42)
<i>orig, ntr30</i> <i>TOP, WML</i>	0.82(0.10)	0.91(0.12) 0.98(0.08)	0.81(0.11) 0.90(0.12)	0.50(0.50) 0.20(0.35)
<i>ss, ntr30</i> <i>TOP, WML</i>	0.84(0.11)	1.00(0.00) 0.90(0.32)	0.85(0.12) 0.96(0.08)	0.63(0.48) 0.30(0.48)
<i>orig, ntr10</i> <i>TOP, NAWM</i>	0.80(0.12)	0.80(0.15) 0.95(0.16)	0.83(0.14) 0.84(0.10)	0.71(0.49) 0.35(0.41)
<i>ss, ntr10</i> <i>TOP, NAWM</i>	0.81(0.10)	1.00(0.00) 1.00(0.00)	0.88(0.11) 0.87(0.19)	0.43(0.37) 0.40(0.52)

Table 7.14: Robustness test three-class problem NC vs. AD vs. LBD. Results for the three-class problem NC vs. AD vs. LBD are reported as mean with standard deviation in brackets, over 10 folds cross validation, classifying NC vs. AD vs. LBD. TotAcc=total accuracy, R=recall, P=precision. 0 for class NC and 1 for class AD and 2 for class LBD. ROI is either *WML* for white matter lesion area or *NAWM* for normal appearing white matter area. *TOP* is short for LBP-TOP. *orig* means that the data used for analysis was resampled and *ss* means that data was from a single scanner only. *ntrXX* refers to the number of trees used in the random forest classifier.

Test	TotAcc	P0 R0	P1 R1
$AD/LBD_{VLBP,WML}^{orig,ntr30}$	0.74(0.13)	0.79(0.08) 0.91(0.12)	0.33(0.52) 0.10(0.21)
$AD/LBD_{VLBP,WML}^{ss,ntr30}$	0.80(0.10)	0.83(0.07) 0.94(0.11)	0.38(0.48) 0.15(0.34)
$AD/LBD_{VLBP,NAWM}^{orig,ntr30}$	0.69(0.12)	0.80(0.09) 0.81(0.15)	0.27(0.36) 0.25(0.35)
$AD/LBD_{VLBP,NAWM}^{ss,ntr30}$	0.78(0.08)	0.81(0.02) 0.94(0.10)	0.25(0.50) 0.05(0.16)
$AD/LBD_{TOP,WML}^{orig,ntr20}$	0.71(0.15)	0.80(0.11) 0.83(0.16)	0.31(0.43) 0.25(0.35)
$AD/LBD_{TOP,WML}^{ss,ntr20}$	0.81(0.12)	0.84(0.11) 0.96(0.10)	0.50(0.50) 0.20(0.42)
$AD/LBD_{TOP,NAWM}^{orig,ntr20}$	0.83(0.17)	0.86(0.12) 0.93(0.12)	0.72(0.44) 0.50(0.41)
$AD/LBD_{TOP,NAWM}^{ss,ntr20}$	0.74(0.11)	0.84(0.09) 0.85(0.14)	0.25(0.27) 0.25(0.42)

Table 7.15: Robustness test two-class problem AD vs. LBD. Results for the two-class problem AD vs. LBD are reported as mean with standard deviation in brackets, over 10 folds cross validation, classifying NC vs. AD vs. LBD. TotAcc=total accuracy, R=recall, P=precision. 0 for class NC and 1 for class AD and 2 for class LBD. ROI is either *WML* for white matter lesion area or *NAWM* for normal appearing white matter area. *TOP* is short for LBP-TOP. *orig* means that the data used for analysis was resampled and *ss* means that data was from a single scanner only. *ntrXX* refers to the number of trees used in the random forest classifier.

7.4.5 Discussion

Results: 3D texture analysis by VLBP/C and LBP-TOP/C shows encouraging results when classifying AD- and LBD dementia and normal controls. The three-class problem NC vs. AD vs. LBD shows a total accuracy of 0.79(0.07), the two-class problems NC vs. AD and NC vs. LBD show a total accuracy of 0.97(0.07) and 0.97(0.06) respectively. The two-class problem AD vs. LBD shows a total accuracy of 0.79(0.15). Compared to the results we achieved in [173] where we used 2D LBP texture analysis, the results using 3D LBP texture features shows lower accuracy for the three-class problem, approximately the same accuracy for the two-class problems NC vs. AD and NC vs. LBD as was achieved in the NC vs. (AD+LBD) problem, and higher accuracy for the two-class problem AD vs. LBD. The latter is especially encouraging since one of the main goals of this work was to test whether 3D texture analysis would improve on that specific task.

In the three-class problem using LBP-TOP/C for 3D texture feature calculation gives higher accuracy as compared to VLBP/C. In the two-class problem NC vs. AD, similar results is achieved for both texture features. VLBP/C gives the best performance for the two-class problems NC vs. LBD and AD vs. LBD. Based on these results it could seem like the VLBP/C feature most often outperforms the LBP-TOP/C than the opposite, but it is not a superior victory. The VLBP/C is more computationally expensive, but this is hardly an argument for choosing one over the other since great computer capacity is easily available today. On the other hand, the VLBP/C approach results in a much lower feature count than LBP-TOP/C which is a benefit combined with many classifiers.

The WML region showed highest accuracy for the NC vs. LBD problem. On the other hand, In the NC vs. AD problem both ROI's showed the same high accuracy and in both the three-class problem and in the AD vs. LBD problem NAWM gave better results than the WML region.

Limitations: A limitation to our study is the variation in number of subjects in the data used for classification and especially the small number of subjects in the LBD group. This can probably explain some of the reason that the AD vs. LBD problem acquires lower accuracy compared to the two other two-class problems. This is the reason we wanted to address the imbalanced data problem by testing two different approaches in addition to doing the classification on the original and untouched data, namely adding cost-sensitivity to the classification and resampling the data such that all classes had same number of subjects as the largest class using SMOTE. Neither in the three-class problem nor the two-class problem AD vs. LBD were there any benefits in adding cost-sensitivity to the classifier in any of the tests. On the other hand, resampling with SMOTE increased classification accuracy to a large extent in the LBD class at the expense of total accuracy and the accuracy of the AD class. In both the two-class problems NC vs. AD and NC vs. LBD, adding cost-sensitivity increased classi-

fication performance in many of the tests, but SMOTE increase accuracy even more in most of the tests. The difficulty in discerning AD from LBD can also be explained by the similarities in the changes appearing in neurodegenerative dementia stated by Bartzokis in [171] and Schneider's observations about mixed brain pathology in [172].

Clinical value: In an attempt to add clinical value to the 3D texture analysis approach, we decided to calculate correlations between the texture features and cognition. Only the AD patients showed a great number of significant correlations and only when the texture features were calculated in the NAWM region. None of the features showed correlation in the LBD group and to a very little extent did the NC subjects show any correlations. The classification results showed highest accuracy in WML as well as NAWM in the NC vs. AD problem. In addition, both the three-class problem as well as the AD vs. LBD problem showed best classification performance when the texture features were calculated in the NAWM region. Disintegration of the normal appearing white matter (NAWM) is strongly related to the severity of WML [174]. This could indicate that WML as well as NAWM is important in dementia research generally and AD especially.

Robustness of methods: The robustness of the proposed approach for dementia classification was addressed by comparing the classification results performed on the original data and data from one scanner only as explained earlier (see Section 7.4.3). We expected that the results from one scanner only data to be similar or slightly better because of higher consistency in the data. This is also what we experienced. Even though we to some degree have been able to develop a robust method for dementia classification, a limitation to our study is that we have not validated the method on another large scale data-set as for example ADNI-data.

Comparison to others: Vemuri et al. [62] diagnosed AD with a sensitivity of 90.7% and a specificity of 84%, LBD with a sensitivity of 78.6% and specificity of 98.8%, and FTLD with a sensitivity of 84.4% and a specificity of 93.8% using atrophy maps and k-means clustering. All the LBD subjects were later histologically confirmed. In the same paper, the sensitivity and specificity for the respective clinical diagnoses were reported as well. AD with a sensitivity of 89.5% and a specificity of 82.1%, LBD with a sensitivity of 70.0% and specificity of 100.0%, and FTLD with a sensitivity of 83.0% and a specificity of 95.6%. In the light of these numbers, our method shows higher accuracy for the AD group and for the LBD group compared to clinical diagnosis. The use of different measures of goodness and that another dataset was used reduces the value of the comparison. In [106] Kodama and Kawase reported a classification accuracy of 70% when classifying LBD patients from AD patients and normal controls. In Burton et al. [61] a sensitivity of 91% and a specificity of 94% is achieved for the AD class in a data material containing subjects with AD, LBD, and vascular

cognitive impairment, using calculations of medial temporal lobe atrophy. Lebedev et al. [65], reports a sensitivity of 94.4 and a specificity of 88.89 discerning AD from LBD using sparse partial least squares (SPLS) classification of cortical thickness measurements from the two groups.

Conclusion: We have been able to develop a CAD system where NC, AD, and LBD are discerned with higher accuracy than clinical diagnosis. 3D texture analysis improved classification accuracy in the AD vs. LBD problem as compared to 2D texture features reported earlier. It did not seem to be important whether the choice fell on VLBP/C compared to LBP-TOP/C as texture features. NAWM seemed to be the ROI that most often gave highest accuracy. Both cost-sensitive classification and resampling using SMOTE proved good methods to handle the imbalanced data problem. The AD group stood out regarding correlation between texture features and cognition, mainly in the NAWM region. Together this could suggest that the NAWM region is a relevant area regarding Alzheimer's disease. The robustness of the method was tested and seems to be good even though a limitation to our study is the lack of validation from for example ADNI data.

7.5 Preliminary results for texture analysis in GM

In the former papers, focus has been put on WML and WM as ROI for the extraction of texture features. The amyloid hypothesis for Alzheimer's disease posted by Hardy and Allsop [175], has been an important basis for work in the field of dementia research for more than two decades. In this work, we wanted to apply local binary pattern (LBP) texture analysis in GM as well as WM to distinguish patients with AD, LBD, and NC with high accuracy and to get an impression whether textural differences between NC-, AD-, and LBD subjects appear mostly in WM or GM. The method used are similar to those used in Paper III using 2D LBP.

A total accuracy of 0.95(0.06) is achieved in the two class problem NC vs. AD+LBD when LBP/C have been calculated using GM as ROI. Correspondingly high values were achieved for precision and recall. The results from the GM and WM areas were very similar and showed high accuracy. This may indicate that there are great differences between healthy elderly persons and patients with neurodegenerative dementia in all regions of the brain which is supported by the results published by Schneider et al. [172] and that dementia may neither be a GM nor a WM disease alone, but that all of the brain is severely affected. In the two class problem AD vs. LBD, the classification performance achieved in the GM region was a total accuracy of 0.72(0.12), slightly worse than the results from the WM region, but still comparable. To conclude – the results from the WM and the GM region are very similar, slightly favoring the WM region.

Chapter 8

Discussion

Throughout the papers included in this thesis, we have performed quantitative analysis of WML as well as performing texture analysis in WML and non-WML regions in subjects with dementia and healthy controls. In the following chapter an overall discussion of the contributions of this thesis, of the limitations that need to be acknowledged as part of this project, and suggestions for future research will be presented.

8.1 Summary of contributions

The contributions given in this thesis span from volumetric quantification of WML in different regions of the brain to applying advanced biomedical imaging techniques, such as texture analysis, in WML- and other brain regions. Conventional statistical methods for group comparison, as well as machine learning techniques have been subject of investigation. Below will a summary of the contributions in this thesis be listed.

8.1.1 Quantitative analysis of WML in dementia

Understanding the role of WML for the pathogenesis of the progression of cognitive impairment is important, since preventing WML may represent a target for future attempts to prevent or slow down the disease process. It is particularly important to study this in the early phase of the disease, as this will likely be the target for future treatment. Few studies have explored the severity and consequences of WML in LBD, the reported findings are inconsistent, and they have been performed in subjects with end-stage disease (autopsy studies) or with moderate to severe dementia [176, 177, 148]. Thus, there has been a need to clarify the influence of WML in mild LBD. A contribution of this thesis has been to measure the total and regional volume of WML in patients with mild AD and LBD and to explore the association with cognitive impairment including

memory and executive function. We concluded that although severity of WML did not differ significantly between subjects with dementia and NC or between mildly demented patients with AD and LBD, WML were associated with cognitive decline in AD, but not LBD, and that more studies of the potential clinical impact of WML in patients with LBD are needed. See Section 2.2 and Paper I for more information.

Several methods for visual quantification of WML severity from MR images of the brain exist [150, 151, 152]. A significant limitation to all of the above referenced methods is that they are prone to inter rater variability. It has been shown that quantification of the actual volume of WML is a more sensitive approach than the visual rating scales [153]. Others have found that visual rating is as good as the more complex methods in routine clinical practice, but that volumetric assessment should be used in research settings if possible [154]. Automated quantitative segmentation methods are suitable for assessing impact of WML on cognitive function [155]. Lately, several approaches for automatic segmentation of WML have been suggested [158, 157, 159]. A contribution of this thesis has been to use automatic segmentation routines for the calculation of total and regional volume of WML. See Section 4.1 and Paper I for more information.

Statistical analysis is a common approach when looking for differences between groups in quantitative analysis in dementia. Lately, the interest in applying ML approaches have increased. In [51], the authors performed classification of patients with AD, DLB and frontotemporal lobar degeneration using results from six visual rating scales as input and report a substantial better performance as compared to using only a single scale as input and at least as good as expert reads. In [178], the authors propose a method for differential diagnosis of AD, frontotemporal lobar degeneration, vascular dementia, DLB, and NC using volumetric- and morphological- as well as vascular characteristics from MRI as input to a multiclass classifier. By combining information from T1 weighted images, DT-MRI, and resting state functional MRI, the authors in [179], were able to classify NC from mild AD from moderate AD with an increasing accuracy by stepwise adding features from the different MRI sequences. Others have had success in predicting development from to AD from MCI with high accuracy by combining quantitative MRI and clinical information such as plasma biomarkers in classification [180]. A contribution of this thesis has been to use WML volume as features in an ML framework classifying subjects with dementia from healthy controls. The best classification results obtained using the WML volumes as features were total cerebellar WML volume with an AUC of 0.73 and 95% confidence interval of 0.57 to 0.83. Similar results were achieved for the cerebellar WML volume in the deep white matter, the frontal WML volume in the periventricular region and the volume of WML in the basal ganglia. See Paper II for more information.

8.1.2 Texture analysis in AD and LBD

AD and LBD are very complex diseases making them difficult to prevent, delay or cure. Current therapy focus on many approaches, for example helping patients maintain an acceptable mental functioning, managing typical behavioral changes, and slowing symptom progression. Early intervention is important, and the ability to identify these types of dementia and healthy controls early in the disease course may be essential for successful patient care. Differentiating between AD and LBD is also important since they differ in prognosis and response to drug treatment. Currently, the only available method to differentiate between AD and LBD is the dopamine transporter scan, which is expensive and not readily available at all centers.

White matter (WM) comprises approximately half the brain volume and provides connectivity between the two brain hemispheres as well as ensure efficient transfer of neural activity complementing information processing in the gray matter (GM). WM neuropathology is often diffuse and affects many neuronal networks which can be disturbed simultaneously resulting in a multidomain syndrome. An important hypothesis for AD has been that the disease is caused by accumulation of peptide amyloid beta ($A\beta$) in the brain. It has been called the amyloid hypothesis and has been a prevailing motivation for research activity for years. Less attention has been brought to studying WM pathology with the ambition of bringing greater understanding of the dementia disease process [181, 171, 182, 174, 34, 33].

CAD can be a helpful tool to pinpoint diagnosis early in the disease course in a cost-effective manner and unbiased to human inconsistencies [76]. Early detection of disease and the discovery of functional connections between brain areas relevant for the disease are important benefits. Recent advances in the field have focused especially on AD and patients with MCI, which are considered a precursor to AD [75, 78, 79, 64, 80]. Less attention has been put into developing CAD systems for LBD. As mentioned above, LBD have high prevalence, and accurate clinical diagnosis depends on little available and expensive dopamine transporter scan and postmortem histology. Few papers report high accuracy discerning patients with AD and LBD or other dementias using neuroimaging techniques on MRI [61, 62, 65].

Not many have applied texture analysis in a machine learning (ML) environment to successfully discern different dementias from each other and from healthy controls [101, 183, 106, 105, 102, 104, 103]. A contribution of this thesis has been to apply 2D- and 3D texture analysis in white matter (WM), WML regions as well as normal appearing white matter (NAWM) on FLAIR and T1-weighted MR images as a computer based application for dementia diagnosis. We experienced better classification results when using texture features calculated from the 3DT1 MR image compared to the FLAIR MR image. A total accuracy, reported as mean(std) over cross validation folds, of 0.97(0.07) or higher was reported for the dementia vs. NC, AD vs. NC, and LBD vs. NC classifica-

tion problems for both the 2D- and 3D texture analysis approaches. In the AD vs. LBD case a total accuracy of 0.73(0.16) was reported using the 2D texture analysis approach slightly exceeded by the 3D texture analysis approach were 0.79(0.15) was reported. See Chapters 4 and 5 and Papers II,III, and IV for more information.

8.1.3 Machine learning in dementia

An important application of pattern recognition is computer aided diagnosis. A common challenge when applying classification to cohort data is the lack of sufficient number of instances in each group. One consequence may be difficulty in learning a classifier with sufficient low bias and variance in the validation results. It is especially important that care is taken when training a classifier on data containing few instances and a high number of descriptors for each instance. This could cause overfitting which refers to a learned classifier that performs well on the training data, but have low generalization abilities resulting in poor validation outcomes. Another consequence of insufficient data is data groups with unequal number of instances, often called the imbalanced data problem. Most learning algorithms will fail to perform accurately under such circumstances. The distributive characteristics of the data is often difficult to describe resulting in poor accuracy across data groups. The induction rules describing the minority group becomes fewer and weaker compared to the majority group. A requirement in classification when the dataset is imbalanced is high accuracy for the minority class without simultaneously reducing the accuracy for the majority class too much. Many approaches have been suggested in the literature in an attempt to remedy the problem [138]. Yet another consequence is that common performance measures such as *accuracy* and *error rate* can possibly give a false impression of classifier performance. A representation of classification performance that takes into account the successful as well as the unsuccessful classification results for both the positive and the negative class in a two class problem is important. A contribution of this thesis has been to construct an ML system for classification of different dementia and healthy controls applying a proper training and testing procedure by using a random forest classifier together with a nested cross validation scheme, handling the imbalanced data problem in a proper manner by using SMOTE and cost sensitive classification, and using proper performance measures such as AUC, precision, and recall for validation of classification results. In Paper III, we concluded that neither in the three-class problem nor the two-class problem AD vs. LBD were there any benefits in adding cost-sensitivity to the classifier in any of the tests. On the other hand, upsampling with SMOTE increased classification accuracy to a large extent in the LBD class at the expense of total accuracy and the accuracy of the AD class. In both the two-class problems NC vs. AD and NC vs. LBD, adding cost-sensitivity increased classification performance in many of the tests, but SMOTE increase accuracy even more in most of the tests. See Chapter 6

and Papers II,III and IV for more information.

8.1.4 ROI-wise analysis

A recurring theme throughout this work has been the ROI-wise analysis in dementia. Firstly as a volumetric analysis of WML in Paper I and II, and then as a region of interest for calculation of texture features. In Paper II texture analysis in the WML region was studied. In Paper III results from 2D texture analysis in WML and all of WM were compared. In Paper IV results from 3D texture analysis in WML were compared to NAWM. Preliminary results on 2D texture analysis in GM have also been presented. It seems like the results do not differ much between regions and varies in an inconsistent way.

8.2 Limitations

A fairly robust system for CAD in dementia using TA has been developed and an acceptable performance is observed, especially when classifying the dementia groups from healthy controls. In the following sections several limitations will be discussed.

8.2.1 Black box system

The LBP texture feature in its original form are typically organized in a histogram for each subject. These histograms can be quite large, especially when the number of neighbors increases and even more when the 3D variants are used. A high dimensional feature vector in combination with a small dataset often results in overtraining when learning a classifier. In an attempt to overcome such issues, reducing the size of the feature vector was necessary. As a first step we decided to calculate statistical measures such as mean, median, standard deviation, skewness and others, of all the LBP values in the region of interest. The benefit was highly discriminative features in our applications. The drawback was that information regarding brain regional importance of features was to some extent lost. We cannot conclude where in the brain the discriminative features are located on a more accurate scale than the choice of ROI. Even though the methods used throughout this thesis have to some extent been used to study regional importance of texture features in dementia, this has been done on a coarse scale. Results have been reported for WML, NAWM, all of WM, and GM (unpublished results). The results in this work do not vary much between brain regions. A consequence is that we do not learn much about where in the brain the distinguishing features are located or what kind of textural changes that distinguishes the two dementia groups from each other and from healthy controls. Such information has a clinical potential and this is a limitation to our studies.

8.2.2 Clinical diagnosis

Even though the accuracy of dementia diagnosis has increased during recent years, a lot of information has to be collected to determine if there is a reversible or irreversible cognitive impairment, the severity of the symptoms, and the cause of the symptoms. The diagnosis involves a complete medical and neuropsychological evaluation of the patient. Examples of tests run at such an evaluation are a review of history or onset of symptoms, medical history and medications, neurological examination, rule out vitamin deficiencies or metabolic conditions, brain imaging, and mental status testing. Many of these tests are prone to subjective evaluation errors, can be time consuming, and may exhibit low- inter- and intra-rater reliability. Thus, clinical dementia diagnosis is not 100% accurate.

8.2.3 Postmortem brain autopsy

The ultimate gold standard for dementia diagnosis is post mortem brain autopsy. A strength of the DemWest study is that such pathological confirmation has been performed on a substantial amount of the participating subjects. Until this date, 46 subjects have undergone brain autopsy by an experienced neuropathologist whom where blind to clinical data. So far, a diagnostic accuracy of approximately 85% of clinical diagnosis of DLB and AD has been confirmed (unpublished data). Out of the 46 subjects with a post mortem brain autopsy, 12 subjects were a part of our studies. Out of those with a mismatch diagnosis only one was part of the studies included in this thesis and that was a subject with a clinical diagnose of DLB and a pathological confirmed AD diagnosis. Our methods did show high classification performance when classifying AD from NC as well as LBD from NC. The results when classifying AD from LBD did not show the same accuracy. The LBD group consisted of only 16 subjects including one subject with a fault diagnosis. In addition, the LBD group was a collection of subjects with both DLB as well as Parkinson disease dementia with the consequence of blurring the differences between disease groups. This may have contributed to the poor statistical results in Paper I and the low classification performance reported in Papers II-IV when classifying the AD group from the LBD group.

8.2.4 Multicenter study

The data used in the work of this thesis is part of the DemWest- and ParkWest studies. The MR images were acquired at many different imaging centers, each center having different MR scanner vendor, -model, or -software version. Even though resources were put into designing a common MR protocol for all centers, deviations from the protocol and differences in equipment reduced the amount of data available for classifier training as well as variability was introduced to the decision applications. Some of these limitations were overcome using cross validation techniques and random forest classification. If used correctly, the former

are able to take care the problems arising using small datasets in classification. The latter is able to handle small datasets, high dimensionality in the feature vector, and variability in the data.

8.2.5 WML segmentation

The chosen WML segmentation method was not robust enough to handle the great variation in MR image quality at hand and time consuming manual editing was necessary to correct for that. Manual editing is also prone to subjectivity and low intra- and inter-correlation coefficient. In future research we would recommend applying a completely automatic WML segmentation procedure such as , [158], [157], or [159] to ensure a reliable segmentation result.

8.3 CAD in clinical practice

A method for automatic dementia diagnosis based on MR imaging could overcome some of the challenges regarding limitations addressed to the clinical diagnose. Not as a substitution for the clinical point of view, but as a second opinion. Either to aid in making decisions whether to further examine, treat, or monitor a person with a subjective impression of reduced cognitive abilities, or to strengthen differential diagnosis and prediction of further cognitive decline. On the other hand, using CAD systems in clinical practice involves several challenges. Very often, these applications demands a certain level of competence in handling advanced computer software and hardware. Indeed does a CAD system demand a certain quality from the data. As mentioned above, two important limitations to our work is related to using data from a multicenter cohort study. One of them is the imbalanced data problem and the other is the lack of image protocol harmonization. To ensure high statistical power and classification performance, it will be beneficial to acquire a balanced dataset with images having limited amount of artifacts and noise for model training. Using harmonized MRI protocols are crucial. Before CAD can be implemented in clinical practice for the diagnosis of individual patients, large, multicenter studies with pre-defined methods of classification are needed.

8.4 Future research

In addition to being computationally efficient and showing high discriminative properties in many applications, the LBP descriptor have many attractive properties such as gray-scale invariance, few parameters, and rotation invariance achieved by simple cyclic shifts. On the other hand, several limitations can be detected, such as limited spatial support, little local textural information, and high sensitivity to noise. Several attempts have been made to remedy some of the limitations [184, 185, 186, 187, 188, 189]. In [78], the authors propose an

extended version of LBP where both intensity-based and difference-based features are developed and show high discrimination ability. These features would be highly interesting to implement in a CAD system of MRI in dementia in the future.

The 3D texture analysis methods used in this thesis were originally developed for dynamic texture recognition tasks. There is reason to believe that 3D texture features applied to 3D MRI could improve the system's ability to discriminate subjects with different dementia types. In [190], the authors propose an extension of the LBP to 3D. A future perspective would be to further study 3D texture features in dementia.

A limitation to our work has been the lack of clinical information in the features used in the classifier. A future objective would be to implement a CAD system using features that provides a higher clinical value, i.e. features that can tell something more specific on what kind of changes are going on in the diseased brain and what kind of differences exists between subjects with different types of dementia. As mentioned in Chapter 2, common approaches to advanced MR neuroimaging in dementia have been to measure the volume of relevant anatomical brain structures in an attempt to measure atrophy [55]. Other approaches have been to analyze cortical thickness [64, 65, 66] or perform shape analysis. Few have looked into texture analysis which has the potential to depict microstructural changes in the brain. Such analysis may be able to aid in diagnosing pathologic conditions, quantifying the severity of pathology, and quantifying temporal changes of a certain pathological condition even before significant atrophy appears. In [191], features with information on 3D texture orientation show ability to depict differences between age matched subject of different gender.

It is difficult to identify precisely the time of onset for the dementia syndrome. The changes going on in the brain probably starts long before the onset of symptoms and many medical interventions are only of help early in the disease course [192]. Future research on neuroimaging in dementia would probably benefit from multiparameter image analysis combining several MRI descriptors, such as volumetry, measurements of cortical thickness, shape analysis, and texture analysis. Including biochemical- and genetic biomarkers together with information on cognitive profile may strengthen analysis even further. By combining analyses from different variants of MRI as well as combining these with results from positron emission tomography (PET), a multimodal approach has made it possible to track the development of brain changes related to AD as it progresses in time [74, 75]. When done in a proper manner, multimodal image analysis has a great potential for achieving further understanding of disease mechanisms and prediction of disease outcome even before symptom development [193, 194].

Bibliography

- [1] M. Prince, A. Wima, M. Guerchet, G.-C. Ali, Y.-T. Wu, and M. Prina. World Alzheimer report 2015 – the global impact of dementia. Alzheimer’s Disease International, August 2015. <https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf>.
- [2] D. P. Perl. Neuropathology of Alzheimer’s disease. *The Mount Sinai Journal of Medicine*, 77(1):32–42, 2010.
- [3] C. Ballard, S. Gauthier, A. Corbett, C. Brayne, D. Aarsland, and E. Jones. Alzheimer’s disease. *The Lancet*, 377(9770):1019–31, 2011.
- [4] P. Scheltens, K. Blennow, M.M.B. Breteler, B. de Strooper, G.B. Frisoni, S. Salloway, and W.M. Van der Flier. Alzheimer’s disease. *The Lancet*, 2016.
- [5] Z. Walker, K.L. Possin, B.F. Boeve, and D. Aarsland. Non-Alzheimer’s dementia 2: Lewy body dementias. *The Lancet*, 386(10004):1683–97, 2015.
- [6] D. Aarsland, A. Rongve, S. P. Nore, R. Skogseth, S. Skulstad, U. Ehrt, D. Hoprekstad, and C. Ballard. Frequency and case identification of dementia with Lewy bodies using the revised consensus criteria. *Dement Geriatr Cogn Disord*, 26(5):445–52, 2008.
- [7] SA Vann Jones and JT O’Brien. The prevalence and incidence of dementia with lewy bodies: a systematic review of population and clinical studies. *Psychological medicine*, 44(04):673–683, 2014.
- [8] D. Aarsland, C. G. Ballard, and G. Halliday. Are Parkinson’s disease with dementia and dementia with Lewy bodies the same entity? *J Geriatr Psychiatry Neurol*, 17(3):137–45, 2004.
- [9] CF Lippa, JE Duda, M Grossman, HI Hurtig, D Aarsland, BF Boeve, DJ Brooks, DW Dickson, B Dubois, M Emre, et al. DLB and PDD boundary issues – diagnosis, treatment, molecular pathology, and biomarkers. *Neurology*, 68(11):812–819, 2007.

- [10] IG McKeith, DW Dickson, J Lowe, M Emre, JT O'Brien, H Feldman, J Cummings, JE Duda, C Lippa, EK Perry, et al. Diagnosis and management of dementia with Lewy bodies: Third report of the DLB consortium. *Neurology*, 65(12):1863–1872, 2005.
- [11] B. Thanvi, N. Lo, and T. Robinson. Vascular parkinsonism – an important cause of parkinsonism in older people. *Age and Ageing*, 34(2):114–9, 2005.
- [12] A. Poggesi, L. Pantoni, D. Inzitari, F. Fazekas, J. Ferro, J. O'Brien, M. Hennerici, P. Scheltens, T. Erkinjuntti, M. Visser, P. Langhorne, H. Chabriat, G. Waldemar, A. Wallin, and A. Wahlund. 2001-2011: A decade of the LADIS (leukoaraiosis and disability) study: What have we learned about white matter changes and small-vessel disease? *Cerebrovascular Diseases*, 32(6):577–588, 2011.
- [13] Leonardo Pantoni. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *The Lancet Neurology*, 9(7):689–701, 2010.
- [14] V. G. Young, G. M. Halliday, and J. J. Kril. Neuropathologic correlates of white matter hyperintensities. *Neurology*, 71(11):804–11, 2008.
- [15] Sarah E Vermeer, William T Longstreth, and Peter J Koudstaal. Silent brain infarcts: a systematic review. *The Lancet Neurology*, 6(7):611–619, 2007.
- [16] Stéphanie Debette, HS Markus, et al. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *Bmj*, 341, 2010.
- [17] Timo Erkinjuntti, Leonardo Pantoni, and Philip Scheltens. Cooperation and networking on white matter disorders: The european task force on age-related white matter changes. *Dementia and geriatric cognitive disorders*, 9(Suppl. 1):44–45, 1998.
- [18] Leonardo Pantoni, Anna Maria Basile, Giovanni Pracucci, Kjell Asplund, Julien Bogousslavsky, Hugues Chabriat, Timo Erkinjuntti, Franz Fazekas, Josée M Ferro, Michael G Hennerici, et al. Impact of age-related cerebral white matter changes on the transition to disability—the LADIS study: rationale, design and methodology. *Neuroepidemiology*, 24(1-2):51–62, 2005.
- [19] LADIS Study Group et al. 2001–2011: a decade of the ladis (leukoaraiosis and disability) study: what have we learned about white matter changes and small-vessel disease? *Cerebrovascular diseases*, 32(6):577–588, 2011.
- [20] Anders Wallin, Tormod Fladby, et al. Do white matter hyperintensities on MRI matter clinically? *BMJ*, 341, 2010.

- [21] Ari Ylikoski, Timo Erkinjuntti, Raili Raininko, Seppo Sarna, Raimo Sulkava, and Reijo 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*, 26(7):1171–1177, 1995.
- [22] Ellen Garde, Erik Lykke Mortensen, Katja Krabbe, Egill Rostrup, and Henrik BW Larsson. Relation between age-related decline in intelligence and cerebral white-matter hyperintensities in healthy octogenarians: a longitudinal study. *The Lancet*, 356(9230):628–634, 2000.
- [23] John Stirling Meyer, Jun Kawamura, and Yasuo Terayama. White matter lesion in the elderly. *Journal of the neurological sciences*, 110(1):1–7, 1992.
- [24] Franz Fazekas, Peter Kapeller, Reinhold Schmidt, Hans Offenbacher, Franz Payer, and Gudrun Fazekas. The relation of cerebral magnetic resonance signal hyperintensities to Alzheimer’s disease. *Journal of the neurological sciences*, 142(1):121–125, 1996.
- [25] R. Barber, P. Scheltens, A. Gholkar, C. Ballard, I. McKeith, P. Ince, R. Perry, and J. O’Brien. White matter lesions on magnetic resonance imaging in dementia with Lewy bodies, Alzheimer’s disease, vascular dementia, and normal aging. *J Neurol Neurosurg Psychiatry*, 67(1):66–72, 1999.
- [26] F. E. de Leeuw, J. C. de Groot, E. Achten, M. Oudkerk, L. M. Ramos, R. Heijboer, A. Hofman, J. Jolles, J. 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. *J Neurol Neurosurg Psychiatry*, 70(1):9–14, 2001.
- [27] Niels D Prins, Ewoud J van Dijk, Tom den Heijer, Sarah E Vermeer, Peter J Koudstaal, Matthijs Oudkerk, Albert Hofman, and Monique MB Breteler. Cerebral white matter lesions and the risk of dementia. *Archives of Neurology*, 61(10):1531–1534, 2004.
- [28] M. Yoshita, E. Fletcher, D. Harvey, M. Ortega, O. Martinez, D. M. Mungas, B. R. Reed, and C. S. DeCarli. Extent and distribution of white matter hyperintensities in normal aging, MCI, and AD. *Neurology*, 67(12):2192–8, 2006.
- [29] H. Baezner, C. Blahak, A. Poggesi, L. Pantoni, D. Inzitari, H. Chabriat, T. Erkinjuntti, F. Fazekas, J. M. Ferro, P. Langhorne, J. O’Brien, P. Scheltens, M. C. Visser, L. O. Wahlund, G. Waldemar, A. Wallin, and M. G. Hennerici. Association of gait and balance disorders with age-related white matter changes: the LADIS study. *Neurology*, 70(12):935–42, 2008.

- [30] H. Soennesyn, K. Oppedal, O. J. Greve, F. Fritze, B. H. Auestad, S. P. Nore, M. K. Beyer, and D. Aarsland. White matter hyperintensities and the course of depressive symptoms in elderly people with mild dementia. *Dementia and Geriatric Cognitive Disorders Extra*, 2:97–111, 2012.
- [31] N. D. Prins, E. J. van Dijk, T. den Heijer, S. E. Vermeer, J. Jolles, P. J. Koudstaal, A. Hofman, and M. M. Breteler. Cerebral small-vessel disease and decline in information processing speed, executive function and memory. *Brain*, 128(Pt 9):2034–41, 2005.
- [32] Kevin S King, Ronald M Peshock, Heidi C Rossetti, Roderick W McColl, Colby R Ayers, Keith M Hulsey, and Sandeep R Das. Effect of normal aging versus hypertension, abnormal body mass index, and diabetes mellitus on white matter hyperintensity volume. *Stroke*, 45(1):255–257, 2014.
- [33] Anil M Tuladhar, Andrew T Reid, Elena Shumskaya, Karlijn F de Laat, Anouk GW van Norden, Ewoud J van Dijk, David G Norris, and Frank-Erik de Leeuw. Relationship between white matter hyperintensities, cortical thickness, and cognition. *Stroke*, pages STROKEAHA–114, 2015.
- [34] Motonobu Fujishima, Norihide Maikusa, Kei Nakamura, Masahiro Nakatsuka, Hiroshi Matsuda, and Kenichi Meguro. Mild cognitive impairment, poor episodic memory, and late-life depression are associated with cerebral cortical thinning and increased white matter hyperintensities. *Frontiers in aging neuroscience*, 6, 2014.
- [35] Ana Verdelho, Sofia Madureira, José M Ferro, Anna-Maria Basile, Hugues Chabriat, Timo Erkinjuntti, Franz Fazekas, Michael Hennerici, John OBrien, Leonardo Pantoni, et al. Differential impact of cerebral white matter changes, diabetes, hypertension and stroke on cognitive performance among non-disabled elderly. the LADIS study. *Journal of Neurology, Neurosurgery & Psychiatry*, 78(12):1325–1330, 2007.
- [36] Joanna M Wardlaw. What is a lacune? *Stroke*, 39(11):2921–2922, 2008.
- [37] Wiesje M van der Flier, Elizabeth CW van Straaten, Frederik Barkhof, Ana Verdelho, Sofia Madureira, Leonardo Pantoni, Domenico Inzitari, Timo Erkinjuntti, Militta Crisby, Gunhild Waldemar, et al. Small vessel disease and general cognitive function in nondisabled elderly the LADIS study. *Stroke*, 36(10):2116–2120, 2005.
- [38] Hanna Jokinen, Hely Kalska, Raija Ylikoski, Sofia Madureira, Ana Verdelho, Wiesje M Van Der Flier, Philip Scheltens, Frederik Barkhof, Marieke C Visser, Franz Fazekas, et al. Longitudinal cognitive decline in subcortical ischemic vascular disease—the LADIS study. *Cerebrovascular diseases*, 27(4):384–391, 2009.

- [39] Wiesje M van der Flier, ECW van Straaten, F Barkhof, JM Ferro, L Pantoni, AM Basile, D Inzitari, Timo Erkinjuntti, LO Wahlund, E Rostrup, et al. Medial temporal lobe atrophy and white matter hyperintensities are associated with mild cognitive deficits in non-disabled elderly people: the LADIS study. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(11):1497–1500, 2005.
- [40] Yong S Shim, Young Chul Youn, Duk L Na, Seong Yoon Kim, H-K Cheong, So Young Moon, Kyung Won Park, Bon D Ku, J-Y Lee, Jee H Jeong, et al. Effects of medial temporal atrophy and white matter hyperintensities on the cognitive functions in patients with alzheimers disease. *European neurology*, 66(2):75–82, 2011.
- [41] Trey Hedden, Elizabeth C Mormino, Rebecca E Amariglio, Alayna P Younger, Aaron P Schultz, J Alex Becker, Randy L Buckner, Keith A Johnson, Reisa A Sperling, and Dorene M Rentz. Cognitive profile of amyloid burden and white matter hyperintensities in cognitively normal older adults. *The Journal of Neuroscience*, 32(46):16233–16242, 2012.
- [42] Reinhold Schmidt, Katja Petrovic, Stefan Ropele, Christian Enzinger, and Franz Fazekas. Progression of leukoaraiosis and cognition. *Stroke*, 38(9):2619–2625, 2007.
- [43] Mohamad Habes, Guray Erus, Jon B Toledo, Tianhao Zhang, Nick Bryan, Lenore J Launer, Yves Rosseel, Deborah Janowitz, Jimit Doshi, Sandra Van der Auwera, et al. White matter hyperintensities and imaging patterns of brain ageing in the general population. *Brain*, page aww008, 2016.
- [44] Sun-Ah Choi, Virgilio Gerald H Evidente, and John N Caviness. Comparing cerebral white matter lesion burdens between parkinsons disease with and without dementia. *Journal of movement disorders*, 3(1):6, 2010.
- [45] G Waldemar, B Dubois, M Emre, J Georges, IG McKeith, M Rossor, P Scheltens, P Tariska, and B Winblad. Recommendations for the diagnosis and management of Alzheimer’s disease and other disorders associated with dementia: Efn guideline. *European Journal of Neurology*, 14(1):e1–e26, 2007.
- [46] J. T. O’Brien. Role of imaging techniques in the diagnosis of dementia. *Br J Radiol*, 80 Spec No 2:71–7, 2007.
- [47] Massimo Filippi and Federica Agosta. Structural and functional network connectivity breakdown in Alzheimer’s disease studied with magnetic resonance imaging techniques. *Journal of Alzheimer’s Disease*, 24(3):455–474, 2011.
- [48] John A Bertelson and Bela Ajtai. Neuroimaging of dementia. *Neurologic Clinics*, 32(1):59–93, 2014.

- [49] Defeng Wang, Steve CN Hui, Lin Shi, Wen-hua Huang, Tianfu Wang, Vincent CT Mok, Winnie CW Chu, and Anil T Ahuja. Application of multimodal MR imaging on studying Alzheimer's disease: A survey. *Current Alzheimer Research*, 10(8):877–892, 2013.
- [50] Paul Malloy, Stephen Correia, Glenn Stebbins, and David H Laidlaw. Neuroimaging of white matter in aging and dementia. *The Clinical Neuropsychologist*, 21(1):73–109, 2007.
- [51] Lorna Harper, Giorgio G Fumagalli, Frederik Barkhof, Philip Scheltens, John T O'Brien, Femke Bouwman, Emma J Burton, Jonathan D Rohrer, Nick C Fox, Gerard R Ridgway, et al. MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. *Brain*, 2016.
- [52] Johannes Pantel, Peter Schönknecht, Marco Essig, and Johannes Schröder. Distribution of cerebral atrophy assessed by magnetic resonance imaging reflects patterns of neuropsychological deficits in Alzheimer's dementia. *Neuroscience letters*, 361(1):17–20, 2004.
- [53] R Duara, DA Loewenstein, E Potter, J Appel, MT Greig, R Urs, Q Shen, A Raj, B Small, W Barker, et al. Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer's disease. *Neurology*, 71(24):1986–1992, 2008.
- [54] Lieke L Smits, Betty M Tijms, Marije R Benedictus, Esther LGE Koedam, Teddy Koene, Ilona EW Reuling, Frederik Barkhof, Philip Scheltens, Yolande AL Pijnenburg, Mike P Wattjes, et al. Regional atrophy is associated with impairment in distinct cognitive domains in Alzheimer's disease. *Alzheimer's & Dementia*, 10(5):S299–S305, 2014.
- [55] Antonio Giorgio and Nicola De Stefano. Clinical use of brain volumetry. *Journal of Magnetic Resonance Imaging*, 37(1):1–14, 2013.
- [56] Giovanni B Frisoni, Nick C Fox, Clifford R Jack, Philip Scheltens, and Paul M Thompson. The clinical use of structural MRI in Alzheimer's disease. *Nature Reviews Neurology*, 6(2):67–77, 2010.
- [57] Philip Scheltens, Nick Fox, Frederik Barkhof, and Charles De Carli. Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion. *The Lancet Neurology*, 1(1):13–21, 2002.
- [58] Lars-Olof Wahlund, Per Julin, Sven-Erik Johansson, and Philip Scheltens. Visual rating and volumetry of the medial temporal lobe on magnetic resonance imaging in dementia: a comparative study. *Journal of Neurology, Neurosurgery & Psychiatry*, 69(5):630–635, 2000.

- [59] Li Su, Andrew M Blamire, Rosie Watson, Jiabao He, Benjamin Aribisala, and John T O'Brien. Tissue microstructural changes in dementia with lewy bodies revealed by quantitative mri. *Journal of neurology*, 262(1):165–172, 2015.
- [60] R Barber, IG McKeith, C Ballard, A Gholkar, and JT O'Brien. A comparison of medial and lateral temporal lobe atrophy in dementia with lewy bodies and alzheimers disease: magnetic resonance imaging volumetric study. *Dementia and geriatric cognitive disorders*, 12(3):198–205, 2001.
- [61] EJ Burton, R Barber, EB Mukaetova-Ladinska, J Robson, RH Perry, E Jaros, RN Kalaria, and JT O'Brien. Medial temporal lobe atrophy on MRI differentiates Alzheimer's disease from dementia with Lewy bodies and vascular cognitive impairment: a prospective study with pathological verification of diagnosis. *Brain*, 132(1):195–203, 2009.
- [62] Prashanthi Vemuri, Gyorgy Simon, Kejal Kantarci, Jennifer L Whitwell, Matthew L Senjem, Scott A Przybelski, Jeffrey L Gunter, Keith A Josephs, David S Knopman, Bradley F Boeve, et al. Antemortem differential diagnosis of dementia pathology using structural MRI: Differential-stand. *NeuroImage*, 55(2):522–531, 2011.
- [63] Giovanni B Frisoni, Martina Bocchetta, Gael Chételat, Gil D Rabinovici, Mony J De Leon, Jeffrey Kaye, Eric M Reiman, Philip Scheltens, Frederik Barkhof, Sandra E Black, et al. Imaging markers for Alzheimer's disease: Which vs. how. *Neurology*, 81(5):487–500, 2013.
- [64] Olivier Querbes, Florent Aubry, Jérémie Pariente, Jean-Albert Lotterie, Jean-François Démonet, Véronique Duret, Michèle Puel, Isabelle Berry, Jean-Claude Fort, Pierre Celsis, et al. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain*, 132(8):2036–2047, 2009.
- [65] A. V. Lebedev, E. Westman, M. K. Beyer, M. G. Kramberger, C. Aguilar, Z. Pirtosek, and D. Aarsland. Multivariate classification of patients with Alzheimer's and dementia with lewy bodies using high-dimensional cortical thickness measurements: an MRI surface-based morphometric study. *Journal of Neurology*, 2012.
- [66] Elijah Mak, Li Su, Guy B Williams, Rosie Watson, Michael J Firbank, Andrew M Blamire, and John T O'Brien. Progressive cortical thinning and subcortical atrophy in dementia with lewy bodies and alzheimer's disease. *Neurobiology of aging*, 36(4):1743–1750, 2015.
- [67] Rosie Watson, Sean J Colloby, Andrew M Blamire, and John T O'Brien. Subcortical volume changes in dementia with lewy bodies and alzheimer's disease. a comparison with healthy aging. *International Psychogeriatrics*, pages 1–8, 2015.

- [68] Rosie Watson, Andrew M Blamire, Sean J Colloby, Josh S Wood, Robert Barber, Jiabao He, and John T O'Brien. Characterizing dementia with Lewy bodies by means of diffusion tensor imaging. *Neurology*, 79(9):906–914, 2012.
- [69] Marco Bozzali and Andrea Cherubini. Diffusion tensor MRI to investigate dementias: a brief review. *Magnetic resonance imaging*, 25(6):969–977, 2007.
- [70] Derek K Jones, Thomas R Knösche, and Robert Turner. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. *Neuroimage*, 73:239–254, 2013.
- [71] Kenneth K Kwong, John W Belliveau, David A Chesler, Inna E Goldberg, Robert M Weisskoff, Brigitte P Poncelet, David N Kennedy, Bernice E Hoppel, Mark S Cohen, and Robert Turner. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12):5675–5679, 1992.
- [72] Michael D Fox, Abraham Z Snyder, Justin L Vincent, Maurizio Corbetta, David C Van Essen, and Marcus E Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9673–9678, 2005.
- [73] JE Galvin, JL Price, Z Yan, JC Morris, and YI Sheline. Resting bold fMRI differentiates dementia with Lewy bodies vs. Alzheimer's disease. *Neurology*, 76(21):1797–1803, 2011.
- [74] Michael Ewers, Giovanni B Frisoni, Stefan J Teipel, Lea T Grinberg, Edson Amaro, Helmut Heinsen, Paul M Thompson, and Harald Hampel. Staging Alzheimer's disease progression with multimodality neuroimaging. *Progress in neurobiology*, 95(4):535–546, 2011.
- [75] Katherine R Gray, Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative, et al. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage*, 65:167–175, 2013.
- [76] John Stoitsis, Ioannis Valavanis, Stavroula G Mougiakakou, Spyretta Golemati, Alexandra Nikita, and Konstantina S Nikita. Computer aided diagnosis based on medical image processing and artificial intelligence methods. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 569(2):591–595, 2006.
- [77] PM Thompson and LG Apostolova. Computational anatomical methods as applied to ageing and dementia. *The British Journal of Radiology*, 2014.

- [78] Manhua Liu, Daoqiang Zhang, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Ensemble sparse classification of Alzheimer's disease. *NeuroImage*, 60(2):1106–1116, 2012.
- [79] Rémi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehéricy, Marie-Odile Habert, Marie Chupin, Habib Benali, Olivier Colliot, Alzheimer's Disease Neuroimaging Initiative, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766–781, 2011.
- [80] S. Kloppel, C. M. Stonnington, J. Barnes, F. Chen, C. Chu, C. D. Good, I. Mader, L. A. Mitchell, A. C. Patel, C. C. Roberts, N. C. Fox, Jr. Jack, C. R., J. Ashburner, and R. S. Frackowiak. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain*, 131(Pt 11):2969–74, 2008.
- [81] G. Alves, B. Muller, K. Herlofson, I. HogenEsch, W. Telstad, D. Aarsland, O. B. Tysnes, and J. P. Larsen. Incidence of Parkinson's disease in norway: the Norwegian ParkWest study. *J Neurol Neurosurg Psychiatry*, 80(8):851–7, 2009.
- [82] V. C. Hachinski, L. D. Iliff, E. Zilhka, G. H. Du Boulay, V. L. McAllister, J. Marshall, R. W. Russell, and L. Symon. Cerebral blood flow in dementia. *Archives of Neurology*, 32(9):632–7, 1975.
- [83] P. K. Molsa, L. Paljarvi, J. O. Rinne, U. K. Rinne, and E. Sako. Validity of clinical diagnosis in dementia: a prospective clinicopathological study. *Journal of Neurology, Neurosurgery, and Psychiatry*, 48(11):1085–90, 1985.
- [84] S. E. O'Bryant, S. C. Waring, C. M. Cullum, J. Hall, L. Lacritz, P. J. Massman, P. J. Lupo, J. S. Reisch, and R. Doody. Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's research consortium study. *Archives of Neurology*, 65(8):1091–5, 2008.
- [85] 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*, 34(7):939–44, 1984.
- [86] F. Fritze, U. Ehrt, H. Sonnesyn, M. Kurz, T. Hortobagyi, S. P. Nore, C. Ballard, and D. Aarsland. Depression in mild dementia: associations with diagnosis, apoe genotype and clinical features. *Int J Geriatr Psychiatry*, 26(10):1054–1061, 2011.

- [87] S. H. Hafsteinsdottir, G. Eiriksdottir, S. Sigurdsson, T. Aspelund, T. B. Harris, L. J. Launer, and V. Gudnason. Brain tissue volumes by apoe genotype and leisure activity-the ages-reykjavik study. *Neurobiology of Aging*, 33(4):829.e1 – 829.e8, 2012.
- [88] M. J. Firbank, A. J. Lloyd, N. Ferrier, and J. T. O’Brien. A volumetric study of MRI signal hyperintensities in late-life depression. *Am J Geriatr Psychiatry*, 12(6):606–12, 2004.
- [89] Maria Petrou and Pedro García Sevilla. *Image processing: dealing with texture*. Wiley, 2006.
- [90] Robert M Haralick, Karthikeyan Shanmugam, and Its’ Hak Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6):610–621, 1973.
- [91] Bela Julesz, EN Gilbert, LA Shepp, and HL Frisch. Inability of humans to discriminate between visual textures that agree in second-order statistics-revisited. *Perception*, 2(4):391–405, 1973.
- [92] M. Tuceryan and A.K. Jain. *The Handbook of Pattern Recognition and Computer Vision*. World Scientific Publishing Co., second edition, 1998.
- [93] Herb Kaizer. A quantification of textures on aerial photographs. Technical report, Boston Univ. Research Laboratoies, Boston, MA, 1955.
- [94] YP Chien and King-Sun Fu. Recognition of x-ray picture patterns. *Systems, Man and Cybernetics, IEEE Transactions on*, (2):145–156, 1974.
- [95] L Harrison. *Clinical Applicability of MRI Texture Analysis*. PhD thesis, University of Tampere, Tampere, September 2011.
- [96] Andrzej Materka, Michal Strzelecki, et al. Texture analysis methods—a review. Technical report, Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels, 1998. http://www.elel1.p.lodz.pl/programy/cost/pdf_1.pdf.
- [97] G Castellano, L Bonilha, LM Li, and F Cendes. Texture analysis of medical images. *Clinical radiology*, 59(12):1061–1069, 2004.
- [98] A. Kassner and R. E. Thornhill. Texture analysis: a review of neurologic MR imaging applications. *American Journal of Neuroradiology*, 31(5):809–16, 2010.
- [99] Mary M Galloway. Texture analysis using gray level run lengths. *Computer graphics and image processing*, 4(2):172–179, 1975.
- [100] Robert M Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.

- [101] P. A. Freeborough and N. C. Fox. MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease. *IEEE Transactions on Medical Imaging*, 17(3):475–9, 1998.
- [102] M. S. de Oliveira, M. L. 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 amnesic mild cognitive impairment and mild Alzheimer's disease. *American Journal of Neuroradiology*, 32(1):60–6, 2011.
- [103] Jing Zhang, Chunshui Yu, Guilian Jiang, Weifang Liu, and Longzheng Tong. 3D texture analysis on MRI images of Alzheimer's disease. *Brain Imaging and Behavior*, 6(1):61–69, 2012.
- [104] T. R. Sivapriya, V. Saravanan, and P. Ranjit Jeba Thangaiah. *Texture Analysis of Brain MRI and Classification with BPN for the Diagnosis of Dementia*, volume 204 of *Communications in Computer and Information Science*, chapter 56, pages 553–563. Springer Berlin Heidelberg, 2011.
- [105] Xin Li, Hong Xia, Zhen Zhou, and Longzheng Tong. 3D texture analysis of hippocampus based on MR images in patients with Alzheimer's disease and mild cognitive impairment. In *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*, volume 1, pages 1–4. IEEE, 2010.
- [106] N. Kodama and Y. Kawase. Computerized method for classification between dementia with Lewy bodies and Alzheimer's disease by use of texture analysis on brain MRI. In *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pages 319–321, 2009.
- [107] V. A. Kovalev, F. Kruggel, H. J. Gertz, and D. Y. von Cramon. Three-dimensional texture analysis of MRI brain datasets. *IEEE Transactions on Medical Imaging*, 20(5):424–33, 2001.
- [108] T. Ojala, M. Pietikainen, and D. Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *International Conference on Pattern Recognition*, volume 1, pages 582–585, 1994.
- [109] Timo Ojala, Matti Pietikainen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.
- [110] D. Unay, A. Ekin, M. Cetin, R. Jasinschi, and A. Ercil. Robustness of local binary patterns in brain MR image analysis. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2098–101, 2007.

- [111] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):915–28, 2007.
- [112] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition*. Academic Press, 2009.
- [113] L. Breiman. Random forests. Technical report, Statistics Department, University of California, Berkely, January 2001. <http://oz.berkeley.edu/users/breiman/randomforest2001.pdf>.
- [114] Sudhir Varma and Richard Simon. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, 7(1):91, 2006.
- [115] Endre Anderssen, Knut Dyrstad, Frank Westad, and Harald Martens. Reducing over-optimism in variable selection by cross-model validation. *Chemometrics and intelligent laboratory systems*, 84(1):69–74, 2006.
- [116] Leo Breiman and Philip Spector. Submodel selection and evaluation in regression. the x-random case. *International statistical review/revue internationale de Statistique*, pages 291–319, 1992.
- [117] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145, 1995.
- [118] Trevor Hastie, Robert Tibshirani, Jerome Friedman, T Hastie, J Friedman, and R Tibshirani. *The elements of statistical learning*, volume 2. Springer, 2009.
- [119] R Bharat Rao, Glenn Fung, and Romer Rosales. On the dangers of cross-validation. an experimental evaluation. In *SDM*, pages 588–596. SIAM, 2008.
- [120] Jeremy MG Taylor, Donna P Ankerst, and Rebecca R Andridge. Validation of biomarker-based risk prediction models. *Clinical Cancer Research*, 14(19):5977–5983, 2008.
- [121] Mark A Hall and Lloyd A Smith. Practical feature subset selection for machine learning. 1998.
- [122] Mark A Hall and Lloyd A Smith. Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper. In *FLAIRS conference*, volume 1999, pages 235–239, 1999.
- [123] JS Urban Hjorth. *Computer intensive statistical methods: Validation, model selection, and bootstrap*. CRC Press, 1993.

- [124] Gavin C Cawley and Nicola LC Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, 11:2079–2107, 2010.
- [125] Damjan Krstajic, Ljubomir J Buturovic, David E Leahy, and Simon Thomas. Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of cheminformatics*, 6(1):1–15, 2014.
- [126] Hongyu Guo and Herna L Viktor. Learning from imbalanced data sets with boosting and data generation: the databoost-im approach. *ACM SIGKDD Explorations Newsletter*, 6(1):30–39, 2004.
- [127] Gary M Weiss. Mining with rarity: a unifying framework. *ACM SIGKDD Explorations Newsletter*, 6(1):7–19, 2004.
- [128] Nitesh V Chawla, Aleksandar Lazarevic, Lawrence O Hall, and Kevin W Bowyer. Smoteboost: Improving prediction of the minority class in boosting. In *Knowledge Discovery in Databases: PKDD 2003*, pages 107–119. Springer, 2003.
- [129] Marcus A Maloof. Learning when data sets are imbalanced and when costs are unequal and unknown. In *ICML-2003 workshop on learning from imbalanced data sets II*, volume 2, pages 1–2, 2003.
- [130] Yanmin Sun, Mohamed S Kamel, Andrew KC Wong, and Yang Wang. Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 40(12):3358–3378, 2007.
- [131] Mahesh V Joshi, Vipin Kumar, and Ramesh C Agarwal. Evaluating boosting algorithms to classify rare classes: Comparison and improvements. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 257–264. IEEE, 2001.
- [132] Foster J Provost, Tom Fawcett, et al. Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. In *KDD*, volume 97, pages 43–48, 1997.
- [133] Foster J Provost, Tom Fawcett, and Ron Kohavi. The case against accuracy estimation for comparing induction algorithms. In *ICML*, volume 98, pages 445–453, 1998.
- [134] Tom Fawcett. Roc graphs: Notes and practical considerations for researchers. *Machine learning*, 31(1):1–38, 2004.
- [135] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

- [136] Tom Fawcett. Using rule sets to maximize roc performance. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 131–138. IEEE, 2001.
- [137] Foster Provost and Pedro Domingos. Well-trained pets: Improving probability estimation trees. 2000.
- [138] Haibo He and Eduardo A Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009.
- [139] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–57, 2002.
- [140] Gustavo EAPA Batista, Ronaldo C Prati, and Maria Carolina Monard. A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explorations Newsletter*, 6(1):20–29, 2004.
- [141] Nathalie Japkowicz and Shaju Stephen. The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449, 2002.
- [142] Gary Weiss and Foster Provost. The effect of class distribution on classifier learning: An empirical study. Technical report, Rutgers University, Department of Computer Science, 2001.
- [143] Jorma Laurikkala. Improving identification of difficult small classes by balancing class distribution. In *Artificial Intelligence in Medicine: 8th Conference on Artificial Intelligence in Medicine in Europe, AIME 2001 Cascais, Portugal, July 1-4, 2001, Proceedings*, volume 2101, page 63. Springer Science & Business Media, 2001.
- [144] Andrew Estabrooks, Taeho Jo, and Nathalie Japkowicz. A multiple re-sampling method for learning from imbalanced data sets. *Computational intelligence*, 20(1):18–36, 2004.
- [145] BX Wang and N Japkowicz. Imbalanced data set learning with synthetic samples. In *Proc. IRIS Machine Learning Workshop*, page 19, 2004.
- [146] Kate McCarthy, Bibi Zabar, and Gary Weiss. Does cost-sensitive learning beat sampling for classifying rare classes? In *Proceedings of the 1st international workshop on Utility-based data mining*, pages 69–77. ACM, 2005.
- [147] Susan Elaine Lomax. *Cost-sensitive decision tree learning using a multi-armed bandit framework*. PhD thesis, University of Salford, 2013.
- [148] E. J. Burton, I. G. McKeith, D. J. Burn, M. J. Firbank, and J. T. O’Brien. Progression of white matter hyperintensities in Alzheimer’s disease, dementia with lewy bodies, and parkinson disease dementia: a comparison with normal aging. *Am J Geriatr Psychiatry*, 14(10):842–9, 2006.

- [149] T. Hattori, S. Orimo, S. Aoki, K. Ito, O. Abe, A. Amano, R. Sato, K. Sakai, and H. Mizusawa. Cognitive status correlates with white matter alteration in parkinson's disease. *Hum Brain Mapp*, 33(3):727–739, 2011.
- [150] F. Fazekas, R. Kleinert, H. Offenbacher, R. Schmidt, G. Kleinert, F. Payer, H. Radner, and H. Lechner. Pathologic correlates of incidental MRI white matter signal hyperintensities. *Neurology*, 43(9):1683–9, 1993.
- [151] P. Scheltens, F. Barkhof, D. Leys, J. P. Pruvo, J. J. Nauta, P. Vermersch, M. Steinling, and J. Valk. A semiquantitative rating scale for the assessment of signal hyperintensities on magnetic resonance imaging. *J Neurol Sci*, 114(1):7–12, 1993.
- [152] L. O. Wahlund, F. Barkhof, F. Fazekas, L. Bronge, M. Augustin, M. Sjogren, A. Wallin, H. Ader, D. Leys, L. Pantoni, F. Pasquier, T. Erkinjuntti, and P. Scheltens. A new rating scale for age-related white matter changes applicable to MRI and ct. *Stroke*, 32(6):1318–22, 2001.
- [153] E. C. van Straaten, F. Fazekas, E. Rostrup, P. Scheltens, R. Schmidt, L. Pantoni, D. Inzitari, G. Waldemar, T. Erkinjuntti, R. Mantyla, L. O. Wahlund, and F. Barkhof. Impact of white matter hyperintensities scoring method on correlations with clinical data: the LADIS study. *Stroke*, 37(3):836–40, 2006.
- [154] A. A. Gouw, W. M. Van der Flier, E. C. van Straaten, F. Barkhof, J. M. Ferro, H. Baezner, L. Pantoni, D. Inzitari, T. Erkinjuntti, L. O. Wahlund, G. Waldemar, R. Schmidt, F. Fazekas, and P. Scheltens. Simple versus complex assessment of white matter hyperintensities in relation to physical performance and cognition: the LADIS study. *J Neurol*, 253(9):1189–96, 2006.
- [155] A. M. Tiehuis, K. L. Vincken, W. P. Mali, L. J. Kappelle, P. Anbeek, A. Algra, and G. J. Biessels. Automated and visual scoring methods of cerebral white matter hyperintensities: relation with age and cognitive function. *Cerebrovasc Dis*, 25(1-2):59–66, 2008.
- [156] M. J. Firbank, J. T. O'Brien, S. Pakrasi, L. Pantoni, M. Simoni, T. Erkinjuntti, A. Wallin, L. O. Wahlund, I. van Straaten, and D. Inzitari. White matter hyperintensities and depression—preliminary results from the LADIS study. *Int J Geriatr Psychiatry*, 20(7):674–9, 2005.
- [157] Paul Schmidt, Christian Gaser, Milan Arsic, Dorothea Buck, Annette Förschler, Achim Berthele, Muna Hoshi, Rüdiger Ilg, Volker J Schmid, Claus Zimmer, et al. An automated tool for detection of flair-hyperintense white-matter lesions in multiple sclerosis. *Neuroimage*, 59(4):3774–3783, 2012.

- [158] Keith M Hulsey, Mohit Gupta, Kevin S King, Ronald M Peshock, Anthony R Whittemore, and Roderick W McColl. Automated quantification of white matter disease extent at 3 t: comparison with volumetric readings. *Journal of magnetic resonance imaging*, 36(2):305–311, 2012.
- [159] S. Damangir, A. Manzouri, K. Oppedal, S. Carlsson, M. J. Firbank, H. Sonnesyn, O. B. Tysnes, J. T. O’Brien, M. K. Beyer, E. Westman, D. Aarsland, L. O. Wahlund, and G. Spulber. Multispectral MRI segmentation of age related white matter changes using a cascade of support vector machines. *Journal of the Neurological Sciences*, 322(1-2):211–6, 2012.
- [160] S. M. Resnick, D. L. Pham, M. A. Kraut, A. B. Zonderman, and C Davatzikos. Longitudinal magnetic resonance imaging studies of older adults: A shrinking brain. *Journal of Neuroscience*, 23:3295–3301, 2003.
- [161] Q. S. Zeng, C. F. Li, Z. Q. Liu, L. Lou, and Y. Cui. Quantitative analysis of brain volume by in vivo magnetic resonance imaging in normal adults. *Acta Academiae Medicinae Sinicae*, 28(6):795–798, 2006.
- [162] M. A. Ystad, A. J. Lundervold, E. Wehling, T. Espeseth, H. Rootwelt, L. T. Westlye, M. Andersson, S. Adolfsdottir, J. T. Geitung, A. M. Fjell, I. Reinvang, and A. Lundervold. Hippocampal volumes are important predictors for memory function in elderly women. *BMC Medical Imaging*, 9(17), 2009.
- [163] T. B. Dyrby, E. Rostrup, W. F. Baare, E. C. van Straaten, F. Barkhof, H. Vrenken, S. Ropele, R. Schmidt, T. Erkinjuntti, L. O. Wahlund, L. Pantoni, 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*, 41(2):335–45, 2008.
- [164] J. Jovicich, S. Czanner, X. Han, D. Salat, A. van der Kouwe, B. Quinn, J. Pacheco, M. Albert, R. Killiany, D. Blacker, P. Maguire, D. Rosas, N. Makris, R. Gollub, A. Dale, B. C. Dickerson, and B. Fischl. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *NeuroImage*, 46(1):177–92, 2009.
- [165] F. Kruggel, J. Turner, and L. T. Muftuler. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. *NeuroImage*, 49(3):2123–33, 2010.
- [166] S. Kloppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, Jr. Jack, C. R., J. Ashburner, and R. S. Frackowiak. Automatic classification of MR scans in Alzheimer’s disease. *Brain*, 131(Pt 3):681–9, 2008.

- [167] Gilberto Sousa Alves, Felipe Kenji Sudo, Carlos Eduardo de Oliveira Alves, Letice Ericeira-Valente, Denise Madeira Moreira, Elias Engelharts, and Jerson Laks. Diffusion tensor imaging studies in vascular disease - a review of the literature. *Dement Neuropsychol*, 6(3):158–63, 2012.
- [168] Martin Dyrba, Michael Ewers, Martin Wegrzyn, Claudia Plant, Annahita Oswald, Michela Pievani, Arun Bokde, Andreas Fellgiebel, Massimo Filippi, Lucrezia Hausner, et al. Predicting prodromal Alzheimer’s disease in people with mild cognitive impairment using multicenter diffusion-tensor imaging data and machine learning algorithms. *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association*, 9(4):426–426, 2013.
- [169] Mala Naik, Arvid Lundervold, Harald Nygaard, and Jonn-Terje Geitung. Diffusion tensor imaging (DTI) in dementia patients with frontal lobe symptoms. *Acta Radiologica*, 51(6):662–668, 2010.
- [170] Michael J Firbank, Andrew M Blamire, Andrew Teodorczuk, Emma Teper, Dipayan Mitra, and John T O’Brien. Diffusion tensor imaging in Alzheimer’s disease and dementia with Lewy bodies. *Psychiatry Research: Neuroimaging*, 194(2):176–183, 2011.
- [171] George Bartzokis. Alzheimer’s disease as homeostatic responses to age-related myelin breakdown. *Neurobiology of Aging*, 32(8):1341–1371, 2011.
- [172] Julie A Schneider, Zoe Arvanitakis, Woojeong Bang, and David A Bennett. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology*, 69(24):2197–2204, 2007.
- [173] Ketil Oppedal, Trygve Eftestøl, Kjersti Egan, Mona K Beyer, and Dag Aarsland. Classifying dementia using local binary patterns from different regions in magnetic resonance images. *International journal of biomedical imaging*, 2015, 2015.
- [174] Susana Muñoz Maniega, Maria C Valdés Hernández, Jonathan D Clayden, Natalie A Royle, Catherine Murray, Zoe Morris, Benjamin S Aribisala, Alan J Gow, John M Starr, Mark E Bastin, et al. White matter hyperintensities and normal-appearing white matter integrity in the aging brain. *Neurobiology of aging*, 2014.
- [175] John Hardy and David Allsop. Amyloid deposition as the central event in the aetiology of Alzheimer’s disease. *Trends in Pharmacological Sciences*, 12:383–388, 1991.
- [176] E. Londos, U. Passant, L. Gustafson, and A. Brun. Neuropathological correlates to clinically defined dementia with lewy bodies. *Int J Geriatr Psychiatry*, 16(7):667–79, 2001.

- [177] E. Ghebremedhin, A. Rosenberger, U. Rub, M. Vuksic, T. Berhe, H. Bickel, R. A. de Vos, D. R. Thal, and T. Deller. Inverse relationship between cerebrovascular lesions and severity of lewy body pathology in patients with lewy body diseases. *J Neuropathol Exp Neurol*, 69(5):442–8, 2010.
- [178] Juha Koikkalainen, Hanneke Rhodius-Meester, Antti Tolonen, Frederik Barkhof, Betty Tijms, Afina W Lemstra, Tong Tong, Ricardo Guerrero, Andreas Schuh, Christian Ledig, et al. Differential diagnosis of neurodegenerative diseases using structural mri data. *NeuroImage: Clinical*, 11:435–449, 2016.
- [179] Tijn M Schouten, Marisa Loitfelder, Frank de Vos, Stephan Seiler, Jeroen van der Grond, Anita Lechner, Anne Hafkemeijer, Christiane Möller, Reinhold Schmidt, Mark de Rooij, et al. Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate alzheimer’s disease. *NeuroImage: Clinical*, 2016.
- [180] Igor O Korolev, Laura L Symonds, Andrea C Bozoki, Alzheimer’s Disease Neuroimaging Initiative, et al. Predicting progression from mild cognitive impairment to alzheimer’s dementia using clinical, mri, and plasma biomarkers via probabilistic pattern classification. *PloS one*, 11(2):e0138866, 2016.
- [181] Faith M Gunning-Dixon, Adam M Brickman, Janice C Cheng, and George S Alexopoulos. Aging of cerebral white matter: a review of MRI findings. *International journal of geriatric psychiatry*, 24(2):109–117, 2009.
- [182] Christopher M Filley. White matter dementia. *Therapeutic advances in neurological disorders*, 5(5):267–277, 2012.
- [183] Tomoharu Kaeriyama, Naoki Kodama, Tetsuo Shimada, and Ichiro Fukumoto. [application of run length matrix to magnetic resonance imaging diagnosis of alzheimer-type dementia]. *Nihon Hoshasen Gijutsu Gakkai zasshi*, 58(11):1502–1508, 2002.
- [184] Shu Liao, Max WK Law, and Albert Chung. Dominant local binary patterns for texture classification. *Image Processing, IEEE Transactions on*, 18(5):1107–1118, 2009.
- [185] Zhenhua Guo, Lei Zhang, and David Zhang. A completed modeling of local binary pattern operator for texture classification. *Image Processing, IEEE Transactions on*, 19(6):1657–1663, 2010.
- [186] Marko Heikkilä, Matti Pietikäinen, and Cordelia Schmid. Description of interest regions with local binary patterns. *Pattern recognition*, 42(3):425–436, 2009.

- [187] Jie Chen, Shiguang Shan, Chu He, Guoying Zhao, Matti Pietikäinen, Xilin Chen, and Wen Gao. Wld: A robust local image descriptor. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1705–1720, 2010.
- [188] Timo Ahonen and Matti Pietikäinen. Soft histograms for local binary patterns. In *Proceedings of the Finnish signal processing symposium, FINSIG*, volume 5, page 1, 2007.
- [189] Xiaoyang Tan and Bill Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Image Processing, IEEE Transactions on*, 19(6):1635–1650, 2010.
- [190] Pedro Morgado, Margarida Silveira, and Jorge S. Marques. Diagnosis of Alzheimer’s disease using 3d local binary patterns. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, 1(1):2–12, 2013.
- [191] V. Kovalev and F. Kruggel. Texture anisotropy of the brain’s white matter as revealed by anatomical MRI. *IEEE Transactions on Medical Imaging*, 26(5):678–85, 2007.
- [192] Mary Bond, G Rogers, J Peters, R Anderson, M Hoyle, A Miners, T Moxham, S Davis, P Thokala, A Wailoo, et al. The effectiveness and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine for the treatment of alzheimers disease (review of technology appraisal no. 111): a systematic review and economic model. 2012.
- [193] Sidong Liu, Weidong Cai, Siqi Liu, Fan Zhang, Michael Fulham, Dagan Feng, Sonia Pujol, and Ron Kikinis. Multimodal neuroimaging computing: a review of the applications in neuropsychiatric disorders. *Brain Informatics*, 2(3):167–180, 2015.
- [194] Sidong Liu, Weidong Cai, Siqi Liu, Fan Zhang, Michael Fulham, Dagan Feng, Sonia Pujol, and Ron Kikinis. Multimodal neuroimaging computing: the workflows, methods, and platforms. *Brain Informatics*, 2(3):181–195, 2015.

Paper I

White matter hyperintensities in mild Lewy body dementia

Authors:

Ketil Oppedal, Dag Aarsland, Michael Firbank, Høgne Sønnesyn, Ole-Bjørn Tysnes, John O'Brien, and Mona K. Beyer

Published in:

Dementia and Geriatric Cognitive Disorders Extra, vol. 2, no. 1, pp. 481-95, 2012.

Available at

<https://www.karger.com/Article/FullText/343480>

Paper II

Using local binary pattern to classify dementia in MRI

Authors:

Ketil Oppedal, Kjersti Engan, Dag Aarsland, Mona K. Beyer, Ole-Bjørn Tysnes,
and Trygve Eftestøl

Printed in:

9th IEEE International Symposium on Biomedical Imaging (ISBI 2012), Barcelona,
pp. 594-597, 2012.

Not available in UiS Brage due to copyright

Paper III

**Classifying dementia using local binary patterns from
different regions in magnetic resonance images**

Authors:

Ketil Oppedal, Trygve Eftestøl, Kjersti Engan, Mona K. Beyer, and Dag Aarsland

Published in:

International Journal of Biomedical Imaging, vol. 2015.

Available at

<http://dx.doi.org/10.1155/2015/572567>

Paper IV

Classifying Alzheimer's disease, Lewy body dementia, and normal controls using 3D texture analysis in magnetic resonance images

Authors:

Ketil Oppedal, Kjersti Engan, Trygve Eftestøl, Mona K. Beyer, and Dag Aarsland

Submitted to a journal 2015

Not available in UiS Brage due to copyright



**University of
Stavanger**

4036 Stavanger
Telefon: 51 83 10 00
Telefaks: 51 83 10 50
E-post: post@uis.no
www.uis.no

ISBN: 978-82-7644-674-6
ISSN: 1890-1387
© 2016 Kjetil Oppedal