Universidade de Lisboa

Faculdade de Ciências

Departamento de Física

U LISBOA | UNIVERSIDADE DE LISBOA

Diagnosis of Alzheimer's disease from 3D images of the brain along time

Filipa da Conceição dos Santos Rodrigues

Dissertação

Mestrado Integrado em Engenharia Biomédica e Biofísica

Perfil em Sinais e Images Médicas

Setembro 2014

Universidade de Lisboa

Faculdade de Ciências

Departamento de Física

Diagnosis of Alzheimer's disease from 3D images of the brain along time

Filipa da Conceição dos Santos Rodrigues

Dissertação

Mestrado Integrado em Engenharia Biomédica e Biofísica

Perfil em Sinais e Images Médicas

Orientador: Prof. Dr. Pedro Almeida
Co-Orientador: Prof. Dr. Margarida Silveira

Setembro 2014

Passion will move men beyond themselves, beyond their shortcomings, beyond their failures.

Joseph Campbell

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Margarida Silveira for giving me the opportunity to work on this interesting project, for all the valuable knowledge and crucial guidance throughout the development of the present thesis. My grateful thanks also go to my advisor Professor Pedro Almeida for his support and encouragement.

I am also thankful to Pedro Morgado for the imaging preprocessing which was a crucial step for the work developed in the present thesis. My thanks also go to Alexandre Ochôa for the helpful discussions.

I would like to thank Maria Inês and Quitério for their friendship and for keeping me sane during this long 5-years journey; and Nanda, for keeping me physically healthy. I am also deeply grateful to Tiago for his never-ending support, optimism and for making me smile everyday.

Finally, I would like to express my deepest gratitude to my parents and brothers, Pedro and Gonçalo, for their love, patience, unconditional support and for the efforts to make all my dreams come true.

Abstract

Alzheimer's disease (AD) is the most common cause of dementia in the elderly. Although no cure has yet been found for this disorder, it is possible to delay progression of its symptoms if therapeutic intervention is provided at the earliest stage of the disease. However, the early diagnosis may be a challenging task for physicians, since the subtle changes in brain tissues associated with the onset of AD are difficult to detect by visual inspection of neuroimaging scans. Hence, in recent years, increasing attention has been given to the development of computer-aided diagnosis (CAD) for AD, in order to assist physicians in image analysis and interpretation. However, the majority of existing CAD systems rely on the analysis of biomarkers at a single time-point, ignoring the progressive nature of the disorder. In the present thesis, the value of incorporating information on cerebral metabolic patterns along time for the automatic classification of AD was investigated. Baseline and multiple follow-up FDG-PET scans of cognitively normal (CN), mild cognitive impairment (MCI) and AD subjects were used. Voxel-based and multi-region analysis of cross-sectional and longitudinal FDG-PET images were performed. In addition, different feature selection methods were tested, as well as several intensity normalization approaches for FDG-PET images. The Support Vector Machine algorithm was used for CN vs AD, CN vs MCI and MCI vs AD classification tasks. Although the longitudinal information did not seem to have a great discriminative power, the combination of longitudinal and cross-sectional data enhanced the classification results achievable using cross-sectional data alone. In fact, this combination led to results that are in line with the current state-of-the-art, suggesting that longitudinal data may provide valuable complementary information for the automatic diagnosis of AD.

Keywords

Alzheimer's Disease, Computer-Aided Diagnosis, [18F] Fluorodeoxyglucose Positron Emission Tomography, Longitudinal Analysis, Intensity Normalization, Feature Selection

Resumo

A doença de Alzheimer (AD) é a forma mais comum de demência e, apesar de atualmente não ter cura, o seu diagnóstico precoce é essencial para agir de forma a retardar a progressão dos sintomas. Por esta razão, o desenvolvimento de sistemas automáticos de diagnóstico usando imagens tridimensionais do cérebro tem despertado um interesse crescente nos últimos anos. No entanto, a maioria dos sistemas propostos baseia-se na análise das imagens apenas num instante temporal, ignorando a natureza progressiva da doença. O presente estudo teve como principal objetivo investigar a relevância da informação sobre o declínio metabólico cerebral ao longo do tempo, para o diagnóstico automático da AD. Foram usados imagens de FDG-PET da baseline e de diversos *follow-up* de sujeitos cognitivamente normais (CN), com défice cognitivo ligeiro (MCI) e pacientes com AD. De forma a extrair as *features* destas imagens, foram utilizadas duas análises distintas, nomeadamente uma abordagem baseada em voxel-a-voxel e outra baseada em regiões de interesse. Além disso, foram testados diferentes métodos para selecionar as *features* mais relevantes, bem como várias abordagens para normalizar a intensidade das imagens de FDG-PET. O algoritmo *Support Vector Machine* foi usado para realizar classificações binárias entre CN vs AD, CN vs MCI e MCI vs AD. Apesar da informação longitudinal demonstrar não ter um grande poder discriminativo por si só, a combinação da informação de um instante temporal com a variação ao longo do tempo levou a melhores desempenhos de classificação, comparativamente a usar apenas os dados num instante temporal. Os resultados obtidos sugerem que a informação longitudinal pode ser um complemento útil para o diagnóstico automático de AD.

Palavras Chave

Doença de Alzheimer, Diagnóstico Assistido por Computador, Tomografia por emissão de positrões com 18F-Fluorodeoxiglucose, Análise Longitudinal, Normalização da Intensidade, Seleção de *Features*

Contents

List of Figures

List of Tables

Acronyms

AD Alzheimer's Disease

ADNI Alzheimer's Disease Neuroimaging Initiative

CAD Computer-Aided Diagnosis

ApoE Apolipoprotein E

APP Amyloid-Precursor Protein

CDR Clinical Dementia Rating

CGM Cerebral Global Mean

CMRgl Cerebral Metabolic Rate of Glucose

CN Cognitively Normal

CSF Cerebrospinal Fluid

FDG Fluorodeoxyglucose

MCI Mild Cognitive Impairment

MMSE Mini-Mental State Examination

MRI Magnetic Resonance Imaging

RBF Radial Basis Function

ROI Region of Interest

SMC Sensorimotor Cortex

SPECT Single Photon Emission Computed Tomography

SPM Statistical Parametric Mapping

SVM Support Vector Machine

PET Positron Emission Tomography

VI Voxel Intensity

1

Introduction

Contents

The aim of the present thesis is to investigate the value of longitudinal information for Computer-Aided Diagnosis (CAD) of Alzheimer's Disease (AD) using 3D brain images. CAD has become a highly researched topic in the medical field in recent years, due to its high potential to improve diagnosis of several types of abnormal conditions based on medical images. Since AD currently has no cure, an early diagnosis is crucial in order to mitigate its symptoms. Hence, CAD schemes may be a valuable tool to assist physicians in image analysis and interpretation, especially at earlier stages of the disease.

This chapter is organized as follows. A detailed description of AD's pathogenesis, epidemiology and diagnosis, with special emphasis on neuroimaging techniques, is presented in Section 1.1. The proposed approach and original contributions are described in Section 1.2 and Section 1.3, respectively. Lastly, in Section 1.4 the outline of the following chapters is provided.

1.1 Motivation - Alzheimer's Disease

1.1.1 Overview

AD is a progressive neurodegenerative disorder and the most common cause of dementia in the elderly [1]. Dementia refers to a state of progressive cognitive decline beyond the expected normal consequence of aging. Since it's a progressive disease, the symptoms gradually worsen over time and dementia severity is usually classified according to the Clinical Dementia Rating (CDR) [2]. This rating is based on the evaluation of 6 different domains: memory, orientation, judgment and problem solving, community affairs, personal care, home and hobbies. CDR has 4 stages in which 0 means no cognitive impairment; 0.5 connotes preclinical AD; 1, 2 and 3 means mild, moderate and severe dementia, respectively. In addition to CDR, another tool is frequently used to evaluate the disease's progression, the Mini-Mental State Examination (MMSE) [3]. This exam aims to assess the overall mental status and consists of 11 scored questions related with memory, attention, orientation, arithmetic and language. The maximum score is 30 and a cutoff is suggested for classification purposes: a score equal to or grater than 27 refers to a Cognitively Normal (CN) subject; a score between 21 and 26 indicates Mild Cognitive Impairment (MCI); a score between 11 and 20 suggests moderate cognitive impairment; a score below 10 reveals a severe case of dementia.

Mild Cognitive Impairment

The MCI stage, also known as preclinical AD, is a transitional state between normal cognitive decline due to aging and dementia. MCI is characterized by an evident memory impairment but with overall preservation of the cognitive function, which can be difficult to diagnose accurately since it may be mistaken for normal aging [4]. Nearly half of the patients diagnosed with MCI progress to dementia within 3 or 4 years, and the majority of these patients declines to AD [5]. For this reason, a distinction between MCI non-converts, i.e, MCI patients who remain stable over a certain period of time, and converts, i.e., MCI patients who will progress to AD in the future has recently received increasing attention in the research field (for example, [6] [7]). The early diagnosis of MCI converts could have a significant impact on the course of dementia, since it would allow an early therapeutic intervention and consequently a delay in the progression of the symptoms.
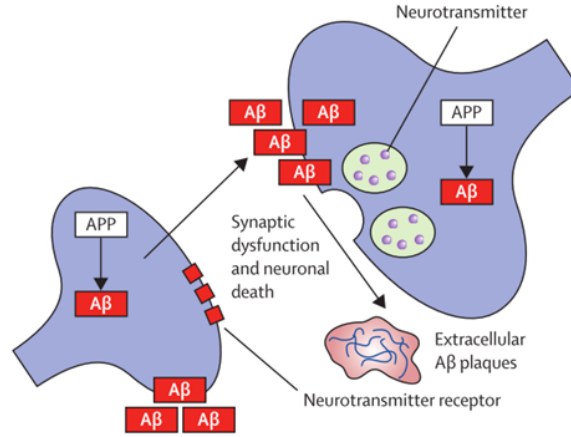
Alzheimer's Disease

The progression of AD is characterized by an overall memory loss and confusion with time and place. The earliest stage of this condition begins with failure of short-term memory and a decline in the ability to perform daily tasks. As the disease evolves, patients start to forget family members' names and personal details such as their address and telephone number. As this stage progresses, cognitive decline worsens and the patient loses the ability to remember recent events, which could lead to behavior and personality changes. At the late stage, AD patients may forget how to perform basic activities and they are completely dependent of help. All verbal abilities and basic motor skills such as walking and swallowing are eventually lost, and the patient becomes unresponsive to the outside world. At the final stage, AD patients are more vulnerable to infections such as pneumonia, which is the most common cause of death [8]. Currently there is no cure for AD and it is an ultimately fatal pathology.

1.1.2 Pathogenesis

AD is characterized by an overall brain atrophy with a significant loss of neurons. However, the exact process underlying the onset of the disease is not completely understood yet. Despite some diverging theories concerning the causes that lead to the onset of AD, there is a number of widely accepted hallmarks. At the microscopic level, the principal hallmarks are the presence of β-amyloid plaques (Aβ) and neurofibrillary tangles, along with neuronal degeneration [1].

Aβ is generated by cleavage of a larger protein called Amyloid-Precursor Protein (APP) which is involved in cell membrane function. The Aβ is present throughout the body with large amounts concentrated in the brain. Under normal conditions, the Aβ present in the brain is degraded and cleared. The central hypothesis for the cause of AD is the failure of this mechanism leading to the accumulation of toxic concentrations of Aβ which aggregates into plaques thus causing neuronal degeneration and consequently leading to dementia - the amyloid cascade hypothesis (Fig 1.1)[9]. The Apolipoprotein E (ApoE) genotype was found to be the gene responsible for the clearance of Aβ [10], hence being associated to a major risk of AD development.

Tangles are composed by tau proteins which have an important role in the nerve cell structure stabilizing the microtubules. Tau proteins are essential to the normal function of the cells since they allow the passage of nutrients. When the microtubule-associated tau protein in neurons becomes abnormally hyperphosphorylated, twisted strands of tau proteins are formed which are accumulated into neurofibrillar tangles. Tangles cause disassembly of the microtubules, disrupting the structure and function of the neuron and consequently the nutrients supply is not ensured, leading to abnormal neuronal and synaptic functions. The presence of these neurofibrillar tangles is used to confirm Alzheimer's pathology in the autopsy ([12], [13]). The tau protein malfunction starts in the earlier stage of the disease affecting the transentorhinal region, spreading in more advanced stages to the hippocampus and amygdala and later to neocortical regions [14].

Figure 1.1: The amyloid cascade hypothesis. APP generates Aβ which is accumulated into plaques, both intra and extracellularly, leading to synaptic dysfunction and neuronal death. Source: Alzheimer's Disease [11].

The amyloid cascade hypothesis suggest that the abnormal function of the tau proteins is trigged by the toxic accumulation of Aβ, although the interaction of these two mechanisms is not clearly understood [15]. Together, Aβ plaques and tau tangles are currently the most accepted hallmarks of AD, even though no consensus exists about which of them plays a more crucial role in the development of AD symptoms.

The most affected brain areas in AD include the hippocampus, amygdala, nuclei basalis, entorhinal cortex and, in advanced stages, high-order association of the temporal, frontal and parietal regions. The lesions' patterns and distribution are usually divided into 6 different stages (I-VI), I being the AD earlier stage and the VI stage the last one (Fig.1.2) [14].

(a) (b) (c)
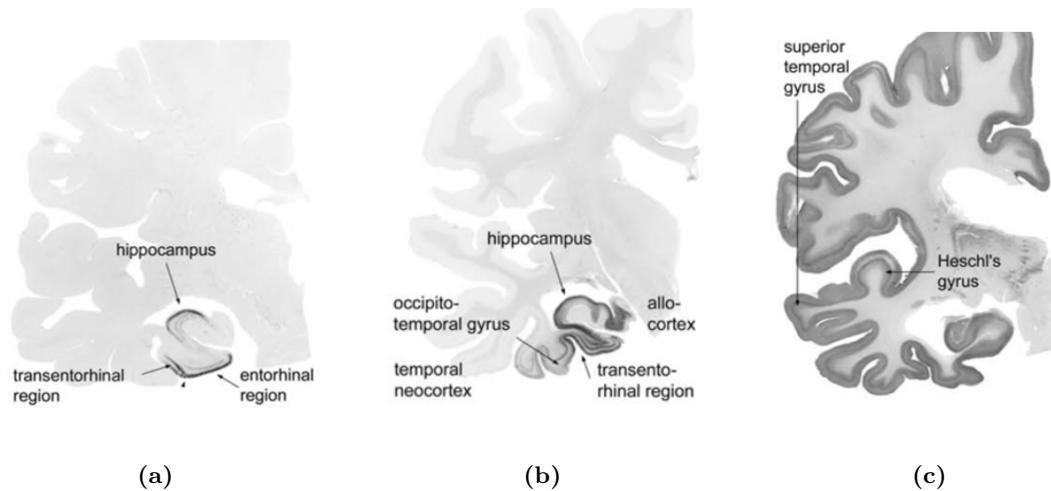
Figure 1.2: Most affected brain regions in AD along different stages. (a) Stage I-II: Transentorhinal region is the first site in cerebral cortex exhibiting alterations in early process of AD. As AD progresses, the lesion extends to the entorhinal region. (b) Stage III-IV: The lesions become more severe and extend into the fusiform and lingual gyri. (c) Stage V-VI: The pathology spreads to the frontal, temporal, occipital, peristriate and striate regions. Adapted from: Braak et al. [14].

1.1.3 Epidemiology

AD was identified for the first time by Alois Alzheimer, a German physician who in 1906 reported the case of a woman who presented unusual symptoms that did not fit into any known disease at the time [16]. Currently, over 35 million people worldwide live with dementia, the majority of whom have Alzheimer's. Since the life expectancy is increasing, due to medicine and technology advances, this number is foreseen to grow in the next years [17]. It is estimated that the number of people living with dementia will reach more than 100 million by 2050, which makes this condition one of the biggest global public health concerns [18].

Although the onset of this disorder is not clearly understood, it is suggested that the risk of developing AD depends on multiple factors rather than a single cause. Age is the greatest risk for AD development, since the majority of the diagnosed cases appear at age 65 or older. There is also evidence that the risk of developing AD tends to increase after this age: the estimated annual incident rate is about 53 new cases per 1000 people aged 65 to 74 years, 75 new cases per 1000 people aged 75 to 84 years and approximately 231 new cases per 1000 to individuals aged 85 years and older [5]. The incidence of AD appears to be greater in women than man (Fig.1.3), which can be explained by women's longer life expectancy. On average AD patients have 4 to 8 years of life after diagnosis, but some can survive as long as 20 years [8].
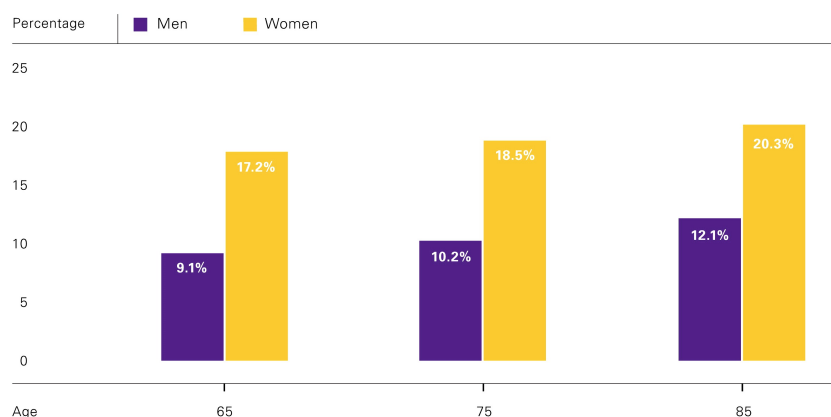
Figure 1.3: Estimated lifetime risk of developing AD by age and sex. Source: Alzheimer's disease facts and figures 2014 [8].

Several studies showed evidence that people with lower level of education seem to have a higher risk to develop AD compared with those with a higher education. Socioeconomic characteristics and quality of life also appeared to be relevant factors, which explains why there is a higher prevalence of AD in countries with a low to middle income (Fig. 1.4).

Severe dementia causes complications which increase the risk of death. The most commonly direct cause of death of AD patients is pneumonia [17]. Even though the correct account for all deaths which were caused primarily by dementia is a difficult task, it was estimated that AD is becoming a more common cause of death. While other common causes of death, such as cancer and HIV conditions have decreased in the past few years, rate of deaths from AD have increased significantly (Fig.1.5) [8].
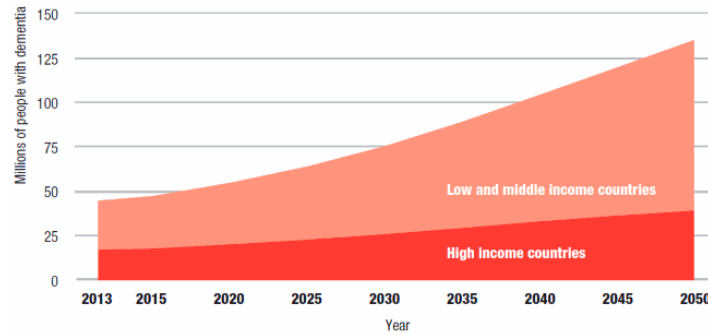
Figure 1.4: Estimated number of people with dementia (millions) until 2050 in high, middle and low income countries. Source: World Alzheimer Report 2010 [18].

Along with the social impact, dementia also has a great economic impact. Three main sectors account for costs associated with dementia: costs of medical care, costs of social care, and indirect costs provided by the patient's family [8]. The estimated annual social cost of dementia worldwide, including only direct costs, is US$604 billion [18].
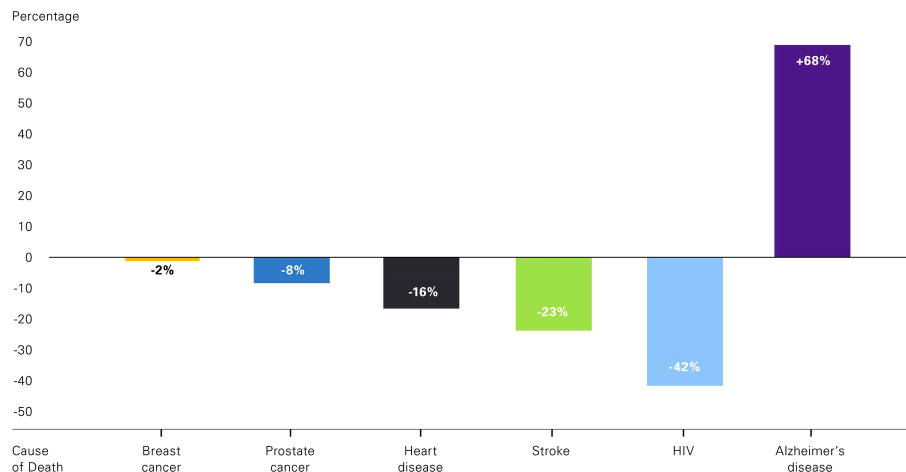
Figure 1.5: Changes (in percentage) in most common causes of death between 2000 and 2010 in United States. Source: Alzheimer's disease facts and figures 2014 [8].

According to the World Alzheimer Report 2010 [17], the estimate of population suffering from dementia in 2010 comprised about 0.5% of the world's total population [17], making this condition a current major public health concern worldwide.

1.1.4 Diagnosis - Neuroimaging Techniques

The primary step for dementia diagnosis consists in the doctor's judgment based on patient's clinical history and reports from the individual, family members and friends. The next step involves cognitive tests and neurological assessment. Although these methods are standard procedures for AD diagnosis, it can only be definitively confirmed in the autopsy. Despite the existence of medications to relieve some of the symptoms and slow down the progression of the disease, currently there is no cure for AD. Therefore, an accurate and early diagnosis is essential in order to delay the progression of the symptoms before they reach severe states.

In 2011, the National Institute on Aging (NIA) and the Alzheimer's Association proposed a new approach for AD diagnosis [19] in order to update the guidelines and criteria used since 1984 [20]. This new approach introduces two considerable changes in the diagnosis criteria: identify three different stages of AD, the earliest one occurring before symptoms (known as preclinical AD or MCI), and incorporate biomarkers tests, such as levels of Aβ and tau protein in the Cerebrospinal Fluid (CSF).

Over the last few decades, metabolic and anatomical studies have been evidencing that some brain regions contribute to early diagnosis of AD, such as the hippocampus and the entorhinal cortex (for example [12], [14]). Neuroimaging techniques, such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET), are widely used for probable AD diagnosis, since they allow a detailed anatomical and physiological examination in a non-invasive manner. A more detailed description of the PET technique, the biomarker used in the present work, is discussed in the following subsections.

FDG-PET

PET is a nuclear imaging technique that uses tracer compounds labeled with positron emitting radioisotopes which are introduced in the body on a biologically active molecule, allowing the measurement of biochemical and physiological processes *in vivo* [21]. Fluorodeoxyglucose (FDG) is the most commonly used tracer. Since brain uses mainly glucose for energy production, labeling glucose with flurorine-18 allows quantitative measurement of the local metabolism [22]. Glucose metabolism has been shown to be closely related with neuronal function and functional activity - higher metabolic activities require higher glucose consumption [23]. Hence, glucose uptake distribution is driven by the neuronal activity and represents brain integrity, while a reduced glucose uptake represents a reduction in number of synapses or a decrease in synaptic metabolic activity [22].

FDG-PET was used for AD diagnosis for the first time approximately 30 years ago, and since then it has been widely applied in research and clinical fields [22]. Several studies found temporal and parietal hypometabolism, as well as other metabolic lesions, in specific regions, such as the anterior cingulate cortex, in patients who suffer from AD (for example [24], [25]). Since MCI condition is a preclinical stage of AD, the AD patients exhibit greater metabolic reductions than the MCI patients (Fig. 1.6). Longitudinal studies have shown the metabolic change to be progressive both in MCI (that may decline to AD) and AD patients (for example [26], [27]). Furthermore, recent studies have shown that AD-like metabolic patterns in MCI patients have higher probabilities to progress to AD within a certain period of time, which suggests that FDG-PET imaging technique is a powerful tool in early diagnosis of AD (for example [22], [28], [29]).
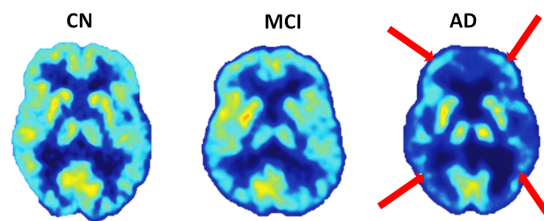
Figure 1.6: Representative examples of brain FDG uptake in CN, MCI and AD subjects. In the AD patient, the arrows indicate the frontal and temporal-parietal regions. Adapted from: G. Small et al. [30].

1.2 Proposed Approach

The majority of existing CAD schemes for AD rely on the analysis of biomarkers at a single time-point. However, since AD is a progressive disorder, changes in biomarkers along time could provide useful complementary information for early diagnosis and progression of AD. The main goal of the present research is to explore the value of incorporating longitudinal imaging data into the classifier to improve the discrimination between AD, MCI and CN subjects. The neuroimaging data used in this work were retrieved from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, including FDG-PET scans from AD, MCI and CN subjects at baseline, 6-month, 12-month and 24-month follow-ups.

First, different methods for FDG-PET intensity normalization were explored. This step is crucial to compare images of different subjects, since the data usually exhibit large intra and inter-individual variation. Intensity normalization is often performed relative to the Cerebral Global Mean (CGM). However, this standard procedure conducts to an attenuation of the differences between clinical groups, which makes the classification task more difficult. Hence, in order to overcome this issue, more suitable approaches are explored, namely the regional mean and the reference cluster intensity normalization methods.

The type of features used play a major role in the success of the classifier to distinguish different clinical groups. Hence, the second intent of this research was to explore two different feature extraction approaches - multi-region and voxel-based analysis. In the multi-region approach, the brain was segmented into different regions, and the mean intensity of each region was extracted. In the voxel-based approach, the intensity of each voxel was used as a feature. Although the first method results in a well balanced number of features and subjects, which avoids a problem that will be further explained, the second one has the advantage of avoiding brain segmentation into regions. However, the pattern of degenerative brain disorders may not follow standard definitions for anatomical or functional regions, therefore, using ROIs may lead to loss of discriminative information.

When the number of features extracted is extremely high compared with the number of subjects available, the *curse of dimensionality* problem arises. This imbalance leads to a performance deterioration of the classifier. In order to deal with this issue, an additional step of feature selection is required so that only the most discriminative features are selected. Furthermore, selecting only the most relevant features leads to a better insight of the physiological process described by the data. Therefore, the advantages of feature selection in the classification task were tested using three different methods: t-test, linear correlation and mutual information.

Finally, in order to investigate the value of the longitudinal information, three different types of datasets were used for the classification: datasets containing only single-time point FDG-PET scans; datasets containing only the changes of the scans over the follow-up period; datasets combining, by simple concatenation, single-time point scans and longitudinal changes.

A Support Vector Machine (SVM) algorithm was used as classifier. Linear and radial basis function kernels were tested in SVM, when the voxel-based and multi-region analysis were performed, respec-

tively. Nested-cross validation procedure was applied in order to tune the classifier parameters and to estimate its performance. Considering the fact that an early diagnosis of dementia is crucial, classification experiments were performed not only to distinguish CN and AD subjects but also to differentiate CN/MCI and MCI/AD individuals.

1.3 Original Contributions

The research in the present thesis brings contributions within the scope of image-based classification of AD and MCI. Recently, some studies have been investigating the value of incorporating longitudinal imaging data for the classification of AD and MCI (for example [31], [32], [33]). Nevertheless, these studies concluded that this information alone did not provide improved classification results compared to using cross-sectional data. Gray et al. in 2012 [33], using a multi-region approach, attempted to combine single-time point scans and changes over a 12-month follow-up period, demonstrating the value of longitudinal information. In the present work, this approach was extended to the entire brain pattern. Hence, a voxel-based approach is proposed, which avoids brain segmentation into regions of interest. Furthermore, the multi-region analysis was also tested, but unlike the work of Gray et al. [33], different feature selection methods were applied in order to select only the most discriminative brain regions for classification. In addition to baseline, other three follow-up FDG-PET scans were explored for classification, namely the 6-month, 12-month and 24-month imaging data.

Regarding the FDG-PET intensity normalization, several studies have been evaluating different methods (for example [34], [35]). The reference cluster method proposed by Yakushev et al. [34], has been widely used since it has proved to increase the discrimination between different clinical groups. In the present thesis, an approach also based on the reference cluster method is proposed.

From the results obtained in the present thesis, a paper entitled "Longitudinal FDG-PET Features for the Classification of Alzheimer's Disease", written in cooperation with my advisor Prof. Margarida Silveira, was submitted and accepted in the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago.

1.4 Thesis Outline

The most important contributions in CAD schemes for AD are summarized in Chapter 2. The role of FDG-PET as biomaker is highlighted as well as the recent growing interest in longitudinal information analysis. The methods used in the present work, including FDG-PET intensity normalization, feature extraction and selection are explained in detail in Chapter 3. A description of the basic concepts and mathematics of SVM is also presented, as well as of the nested-cross validation procedure for parameter optimization and the classification performance assessment. The experiments and results obtained are described in Chapter 4. Finally, Chapter 5 concludes the present thesis, discussing the results, the main research contributions and future work.

2

State of the Art

The uncertainty and subjectivity underlying medical image analysis and interpretation led to the development of computerized systems aiming to assist clinicians. CAD schemes have been widely used as a complement to the physician's diagnosis in order to improve accuracy performance, mainly in cancer detection [36]. The success of this implementation led to a growing interest in expanding CAD schemes to other medical fields. In the last few years, several research programs have focused on the development of these systems to distinguish different types of dementia and to diagnose MCI, AD and other neurological disorders such as Parkinson disease, Lewy body dementia and frontotemporal degeneration [37].

The construction of a CAD scheme is usually based on a supervised learning strategy, i.e., an hypothesis is learned from labeled data, and the label of a new unseen example is determined. The labeled data consists on a set of diagnosed samples, e.g., a set of medical images from healthy subjects and from patients diagnosed with a certain pathology. The labeled data used to train the model is called training data set, and the set where the constructed model is validated is called test (or validation) set. CAD systems involve feature extraction from images used as training data and construction of a predictive model with a machine learning algorithm capable of classifying a new individual image as healthy or pathological. These systems can be constructed using only image-based information or in combination with other relevant biomarkers. The most common methods involved in the construction of CAD systems for AD will now be briefly described (a more detailed explanation of some of these procedures will be presented in Chapter 3).

A biomarker is a biological feature measured *in vivo* that indicates the presence or absence of a certain pathology. Since the criteria for AD diagnosis was established for the first time in 1984 [20], significant progress has been made in the identification of AD-associated brain changes and the incorporation of new biomarkers has been proposed [19]. Current biomarkers for AD include biochemical, neuroanatomical, metabolic, genetic and neurophysiological features. MRI and FDG-PET are widely used neuroimaging techniques in several proposed CAD schemes for AD. FDG-PET proved to be very sensitive to synaptic dysfunction and hypometabolism revealing the presence of neuropathologies before the atrophy of certain structures of the brain detected with MRI technique [22]. Nevertheless, MRI is also widely used in CAD systems for AD, since it has been shown to be very efficient in distinguishing AD subjects from healthy controls or patients with other neurological disorders [38]. Other neuroimaging techniques such as Single Photon Emission Computed Tomography (SPECT) is also explored in various studies (for example, [39], [40]). In addition to neuroimaging data for AD automatic diagnosis, the incorporation of other biomarkers including CSF levels of tau protein, which reflects the brain increase deposition of Aβ, and genetic factors, such as the presence of *APOE* genotype, was also suggested in several studies (for example, [31], [32]).

Feature extraction is a crucial step for CAD schemes since it implies extracting relevant information from medical images, which have great influence on the success of the classifier to identify pathological conditions. Different feature extraction methods have been proposed but the most simple and direct method is the extraction of the Voxel Intensity (VI) values from the whole brain or from a Region of Interest (ROI). Usually, a problem arises with the whole brain approach due to the number of features

highly exceeding the number of samples, the so called *curse of dimensionality*. Thus, a feature selection step is usually implemented in order to ascertain which of the features are relevant for the classification task. Several feature selection methods have been proposed in recent years. Filter methods are one of the most simple and most used approaches. These methods assess the relevance of each feature, in order to decide whether it should be incorporated in the learning stage. Mutual information, correlation coefficients and t-test statistics are some examples of this class of algorithms used as feature selection methods in CAD systems proposed for AD. A detailed description of the aforementioned approaches will be presented in Chapter 3.3.

Regarding the classification task, the majority of the classification methods found in the literature are based on SVM. SVM is a supervised learning algorithm for binary classification, i.e., given a training data classify new unseen examples into one of two classes. A detailed explanation of this algorithm will be addressed in Chapter 3.4.

A chronological review of the most relevant contributions for automatic diagnosis of AD, including a brief description of the methods and principal results obtained, is now presented.

In the past few years, several CAD schemes for AD have been proposed using different neuroimaging techniques. The beginning of the new millennium marked a new era in this field, since before that the majority of the research only aimed to study brain differences between AD and CN subjects, not providing a tool capable of classifying, in an automatic manner, a new individual image. In 2002, Herholz et al. [25] conducted one of the largest multi-center studies accomplished so far. The aim was the development of a fully automated method for FDG-PET images analysis in order to discriminate between CN and AD patients based on a Statistical Parametric Mapping (SPM) approach. An age correction was performed in the measured FDG uptake in each image and t-maps were calculated between CN and AD clinical groups. A score was computed for each individual image as the voxel-by-voxel sum of all t-values of the voxels with FDG uptake below 95% age-adjusted prediction in predefined areas typically affected by AD. Hence, a score higher than an established threshold indicates an abnormal metabolism. This score was shown to be highly susceptible to scan abnormalities, reporting 93.0% of sensitivity and specificity and 97.0% of accuracy to distinguish CN from AD subjects. Although the method proposed by Herholz et al. led to very optimistic results and was adopted in several other later studies, this approach requires *a priori* knowledge about the disease process which may be a drawback. In 2005, Stoeckel et al., [40] proposed an automatic classification method, based on SPECT imaging data, that avoided explicit knowledge about the disease patterns. Rather than simply selecting the most relevant voxels for classification, as the same group has done in a previous work [41], they also incorporated spatial information, generating a classifier that selected the most relevant areas giving a better insight into the physiological patterns of AD. Thus, a contiguous-SVM was proposed as classifier where the model depended on clusters of voxels rather than on isolated ones. They also compared these results with the sensitivity and specificity of the diagnosis based on visual analysis of the SPECT images by expert physicians. The proposed automatic method achieved a sensitivity of 84.4% and 90.0% of specificity, which outperformed the human analysis.

Since AD is a progressive neurodegenerative disorder, discrimination of different stages of the disease has received increasing attention and MCI patients have been included in several studies. In some of these studies, which will be referred to later, a distinction between MCI patients that declined to AD (MCI converters), and MCI patients that remained stable over a certain period of time (MCI non-converters) was also incorporated. In the work of Davatzikos et al. [42] in 2006, a high-dimensional MRI image analysis and pattern classification method was proposed to identify MCI subjects. The MRI images were segmented into gray matter, white matter and CSF. The Pearson-correlation was calculated between each voxel and the class label in order to estimate the specific patterns of MCI condition. In order to retain spatially-meaningful regions, among the features with best values of this coefficient only the ones with spatial consistency with its neighborhood were used as features. A SVM recursive feature elimination was also applied in order to discard less relevant features, achieving an accuracy of 90.0% in the discrimination of CN and MCI subjects. In this work, Davatzikos et al. demonstrated that the subtle structural patterns that characterize the brain of individuals suffering from MCI can be identified from MR scans via high-dimensional analysis and pattern classification methods.

Other studies, in an attempt to gather more meaningful information from a physiological point of view, selected specific brain regions for classification rather than the whole brain. These regions are tipically brain structures or areas known to be affected in AD, such as the hippocampus. In 2008, Colliot et al. [43] developed an automated method for hippocampus segmentation on MR images, and compared this method with manual segmentation. Hippocampal volume differences between groups were assessed using t-test statistics. For classification, a bootstrap method was applied, which consisted on using approximately 75% of the imaging data of each group to obtain a training set and estimate the mean of hippocampal volume. The remaining 25% were used as test set, and the procedure was repeated 5000 times. Each case of the test set was assigned with the closest group, i.e, the group that had the hippocampal volume mean closer to the example being tested. This procedure led to an accuracy of 84.0% classifying AD from CN subjects, an accuracy of 73.0% classifying MCI from CN subjects and an accuracy of 69.0% classifying MCI from AD subjects. Although the hippocampus is one of the most affected brain structures at earlier stages of AD, there are other regions which are affected as the disease progresses. Hence, some studies focused on finding the most affected regions using different types of neuroimaging techniques in order to use them for classification. Magnin et al., in 2009 [44] proposed an automated method to discriminate AD and CN subjects, using a whole-brain segmentation approach from MR images. The whole brain was segmented into 90 anatomical regions, and a t-test was conducted to find the most discriminative regions. An accuracy of 94.5% (91.5% of sensitivity and 96.6% of specificity) was obtained using this approach.

As one can observe, some methods suggested in the aforementioned studies seem to have more success than others. However, this comparison is very difficult to assess since the database used by different studies is usually not the same. Concerned by this observation, Cuingnet et al., in 2011 [45], compared the performance of several different high-dimensional classification methods on the same dataset using 10 different approaches, including whole brain and ROI methods. SVM was used as the

learning algorithm since it is the most commonly used classifier. They concluded that for distinguishing AD from CN, whole brain methods had more success, yielding a higher accuracy. On the other hand, for classifying MCI from CN, hippocampus based methods remained closer to whole-brain methods results. This outcome suggest considering the whole brain may be a more advantageous approach at more advanced stages of the disease.

Currently, a multi-modality approach is emerging. Noticing that different biomarkers provide complementary information valuable to distinguish AD, MCI and healthy individuals, attracted great interest to the investigation of multiple biormakers combination. In 2011, Zhang et al. [32] combined MRI, FDG-PET and CSF biomarkers to discriminate between AD, MCI and CN subjects. For each MRI and FDG-PET image, 93 features were extracted corresponding to 93 anatomical brain segmented regions, and linear SVM was used as classifier. An accuracy of 93.2% was achieved for classifying AD from CN subjects when combining all the three modalities, and only 86.5% of accuracy was achieved when using the best individual modality alone. To distinguish MCI from CN individuals, a classification accuracy of 76.4% was obtained against 72.0% of accuracy when using only the best modality. In this study, a distinction between MCI subjects who convert to AD within 18 months and MCI subjects who remained stable during the same period of time was also considered, and 91.5% of MCI converts and 73.4% of non-converts were correctly classified using the multi-modality approach. The results obtained in this study highlighted the value of combining different biomarkers for the automatic detection of MCI and AD conditions.

More recently, the potential of incorporating longitudinal information for classification has been evaluated. Chen et al. [27] compared the decline of brain metabolic rate over 12 months in AD, MCI and CN patients, using FDG-PET imaging data. In their study, the most discriminative ROIs between groups, i.e., the voxels of the regions consistently associated with longitudinal change, were used. The ROIs were defined using independent two-sample t-test to evaluate which were the more discriminative regions between subjects of each group. In their longitudinal analysis, significant group differences between AD, MCI and CN subjects were reported, which encouraged further longitudinal analysis. In 2011, Hinrichs et al. [31] suggested a multi-modality approach, incorporating cross-sectional and longitudinal MRI and FDG-PET imaging data, clinical measures (CSF and APOE genotype) and neurophysiological status information, using kernel combination methods. As in the work of Zhang et al. [32], they also reported a higher accuracy for multi-modality approach, compared with using any individual biomarker. For their longitudinal analyses, a baseline and a 24-month follow-up MRI and FDG-PET scans were included. They observed that longitudinal analysis of the FDG-PET images had poor discriminative ability, since neither of the two methods considered (voxel-wise temporal differences and voxel-wise temporal ratio) had an accuracy higher than 65.0%. These results suggested that the changes over a 2-year period alone do not provide sufficient information to identify AD with accuracy. Actually, due to this poor performance, they decided not to include the longitudinal information in their final classifier. In 2012, Gray et al. [33] also explored the value of combining cross-sectional and longitudinal multi-region information for classification. Whole-brain segmentation into 83 regions was performed for baseline and 12-month FDG-PET images. The signal intensity of the baseline and

12-month images and signal intensity changes over 12 months for the 83 regions were used as features. Classification accuracies of 88.0% between AD and healthy controls, 81.3% between CN and MCI converters, 83.5% between CN and MCI non-converters and 63.1% between MCI converts and MCI non-converts were achieved by combining cross-sectional and longitudinal information. These results were significantly higher than the ones obtained using cross-sectional data alone, which suggests that longitudinal information may provide valuable complementary information and consequently improve the classification performance.

The current tendency towards a multi-modality combination approach raised an important question related to feature selection. This step is usually performed separately for each modality which ignores the inter-modality relation that gives different yet complementary information. To address this issue, Liu et al., in 2013 [46] proposed a novel approach to multimodality feature selection. The aim was to preserve the complementary information and the relationship between the features derived from different biomarkers. The brain was segmented into 93 ROIs, and gray matter volume of each ROI was extracted as the features for MRI modality and the averaged intensity of each ROI was used as the features for PET modality. The method consisted firstly in treating features from different modalities as different tasks, and then a constraint was imposed in order to preserve the inter-modality relationship. For combining the selected features, a multi-kernel SVM was used. The key of multi-task learning approach relies on the capture of intrinsic relationship between different tasks, exploiting the commonalities between them. An accuracy of 94.3% was achieved for classifying AD from CN subjects, and an accuracy of 78.8% for classifying MCI from CN subjects. A classification beetween MCI converters and non-converters was also assessed, yielding an accuracy of 70.0%, outperforming the results obtained by using other state-of-the-art methods.

A summary of the aforementioned studies is present in table 2.1 .

Author(s), Year	Biomarker(s)	Methods	Participants			*Results (%)			Main Purpose
			CN	AD	MCI	ACC	SENS	SPEC	
Herholz et al., 2002 [25]	FDG-PET	SPM	110	395	-	97.0^1	93.0^1	93.0^1	Automatic method to detect abnormal brain metabolism
Stoeckel et al., 2005 [40]	SPECT	Contiguous SVM	99	31	-	-	84.4^1	90.0^1	Incorporation of spatial information about the features into the classifier
Davatzikos et al., 2006 [42]	MRI	Pearson Corr. Coef., SVM	15	-	15	90^2	-	-	High dimensional image analysis and pattern classification methods to identify MCI subjects
Colliot et al., 2008 [43]	MRI	Hippocampus segmentation, Bootstrap	25	25	24	84.0^1 73.0^2 69.0^3	84.0^1 75.0^2 67.0^3	84.0^1 70.0^2 71.0^3	Automatic segmentation of the hippocampus on MR images
Magnin et al., 2009 [44]	MRI	Whole-brain segmentation, SVM	22	16	-	94.5^1	91.5^1	96.6^1	Whole-brain anatomical segmentation of MR images
Cuingnet et al., 2010 [45]	MRI	Whole-brain, hippocampus segmentation, SVM	162	137	213	-	81.0^1 73.0^2	95.0^1 74.0^2	Comparison of different classification methods
Zhang et al., 2011 [32]	FDG-PET MRI CSF	ROI, SVM	52	51	99	93.2^1 76.4^2	93.0^1 81.8^2	93.3^1 66.0^2	Combination of multiple biomarkers for classification
Hinrichs et al., 2011 [31]	FDG-PET MRI CSF APOE	Multi-kernel learning	66	48	119	92.4^1	86.7^1	96.6^1	Combination of multiple biomarkers and multi-kernel learning for classification
Gray et al., 2012 [33]	FDG-PET	ROI, SVM	54	50	117	88.0^1 81.3^2 83.5^3 63.1^4	83.2^1 79.8^2 79.9^3 52.2^4	96.3^1 82.9^2 86.4^3 73.2^4	Combination of cross-sectional and longitudinal information for classification
Liu et al., 2013 [46]	PET MRI	Multi-kernel, SVM	52	51	99	94.4^1 78.8^2 67.8^4	94.7^1 84.9^2 64.9^4	94.0^1 67.1^2 70.0^4	Multi-task feature selection to preserve inter-modality information

Table 2.1: Chronological summary of some proposed CAD systems for AD since 2002. Biomarkers, method and participants used by each study are present, along with the main results obtained (ACC-accuracy; SENS-sensitivity; SPEC-specificity). The last column resumes the main goals/contributions of each study.
*Results present are the best results achieved in each study. The superscript numbers follows the scheme:
1 - CN vs AD; 2 - CN vs MCI; 3 - MCI vs AD; 4 - MCI converters vs MCI non-converters.

3

Proposed Methods for Classifying CN/MCI/AD

Contents

In the present chapter, the fundamental steps required for the construction of the proposed CAD scheme are described, which include the FDG-PET intensity normalization, feature extraction, feature selection and classification. Intensity normalization of the FDG-PET images is an important preprocessing step. In order to elect the most suitable normalization method for the study population, different approaches were tested and a description of these methods is presented in Section 3.1. The next step involved in the construction of the CAD scheme was the extraction of features from the FDG-PET images. Two approaches were investigated in the present work: multi-region and voxel-based analysis. They are both described in Section 3.2 for cross-sectional and longitudinal data. Regarding the feature selection step, three different methods were tested, namely the correlation coefficient, the t-test and the mutual information. A detailed description of these methods is presented in Section 3.3. Finally, the classification task was accomplished with a SVM algorithm whose basic concepts and mathematics are introduced in Section 3.4.

3.1 FDG-PET Intensity Normalization

The Cerebral Metabolic Rate of Glucose (CMRgl) of cognitively healthy individuals and patients suffering from a neurological disorder exhibit large inter and intra-individual variation. Hence, intensity normalization of FDG-PET images is a crucial preprocessing step in order to allow direct comparison of the data. Several studies (for example, [35], [47], [48]) investigated the effect of using different intensity normalization methods, reporting significant impact in the classification results. Three main approaches have been applied, namely the CGM, the regional mean and the reference cluster intensity normalization. The most commonly used method is the CGM intensity normalization, which uses the whole brain to normalize the images, contrary to the regional mean intensity normalization approach which normalizes the data using only a specific brain region. More recently, a reference cluster approach has been proposed, which is based on a data-driven procedure.

The following subsections present a detailed description of the aforementioned methods.

3.1.1 Cerebral Global Mean Intensity Normalization

Ratio normalization of the regional FDG uptake relative to the CGM has been the most frequently used method. It consists on the division of the intensity value of each voxel by the mean intensity of all intracerebral voxels. When applying this procedure, there is the fundamental requirement that the cerebral global mean does not vary significantly between the different clinical groups. However, this imposition is usually violated in neurodegenarative disorders, such as AD, which have lower CMRgl relative to healthy subjects since the former have the normal neuronal function compromised. Thus, this process conducts to an attenuation of the differences between clinical groups, since the intensity signal of FDG-PET images from patients are artificially scaled up while those from healthy individuals are scaled down. Therefore, this type of normalization leads to an apparent hypometabolism in healthy subjects in regions that are known to be relatively spared in AD.

3.1.2 Regional Mean Intensity Normalization

Recent studies suggest that using specific brain regions rather than the CGM for normalization avoids the bias introduced by the CGM method, leading to an improvement of clinical groups discrimination (for example, [24], [34], [35]). The selected regions are brain areas identified by several studies (for example, [25], [49]) as the most preserved by the disease process, such as cerebellum, brainstem, basal ganglia and Sensorimotor Cortex (SMC). However, these methods require the selection *a priori* of one of these spared areas, based on the assumption that one of these brain regions is the most appropriate to normalize a certain study population.

### 3.1.3 Reference Cluster Intensity Normalization

In contrast with the previous methods, a data-driven approach based on an iterative procedure to define the region used for normalization *a posteriori* was originally proposed in 1997 by Andersson [50]. This method aimed to avoid the bias introduced by the CGM approach when performing PET activation studies. Additionally, the chosen region depends directly on the data being analyzed which also avoids the *a priori* definition of a region. More recently, Yakushev et al., in 2009 [34] proposed a similar data-driven method, but unlike the original algorithm [50] that uses multiple iterations, this method uses only two iterations. In the first one, the standard CGM normalization is performed. As previously explained (Section 3.1.1), this procedure leads to the appearance of hypermetabolic voxels in the pathological group compared to the healthy control in preserved brain regions. In the second iteration, a t-test is conducted in order to find the apparently hypermetabolic regions in the patient group compared to the healthy one. The result of this analysis is expressed by a SPM map, where each voxel is defined by a t-statistic. A threshold for the t-value and for the spatial extent of contiguous voxels is applied. The voxels which fulfill the imposed constrains constitute the clusters and the cluster containing the voxel with the highest t-value is used as reference cluster. The mean value of the voxels of the reference cluster is then extracted and the voxels of each individual image are multiplied by the mean value of the corresponding cluster-derived. Hence, this method selects only the most preserved brain area in a study population. Due to its clear advantages and positive influence on the performance results, this method has been widely adopted in several recent studies aiming to discriminate CN from AD and MCI subjects (for example, [33], [51]) and also from other neurodegenerative disorders (for example, Frontotemporal Lobar Degeneration [52], Parkinson's Disease [53]).

3.2 Feature Extraction

The feature extraction step is a critical procedure that transforms the input data into vectors that must contain all of the relevant information regarding the detection of pathological conditions. Since the original neuroimaging data is extremely high-dimensional but generally a small sample size is available, the feature extraction can also be applied to reduce the amount of input data in order to overcome the *curse of dimensionality* (see Section 3.3). A commonly used method to reduce the input features dimensionality is to group voxels into multiple anatomical regions (multi-region) rather than to use single voxels as features (voxel-based). In the present thesis, a multi-region and a voxel-based analysis were explored for extracting the features from the cross-sectional and longitudinal data.

3.2.1 Cross-sectional Features

Regarding the feature selection for FDG-PET data, the VI-based analysis is the most simple and direct method used. The VI in the FDG-PET modality represents the CMRgl, which is related to neuronal function (see Section 1.1.4). Two main approaches can been applied: extraction of the VI values from the whole brain (voxel-based analysis) or from one or multiple ROIs (multi-region analysis). In the voxel-based analysis, intracerebral VI of the whole brain are extracted and used as input features for the classifier (Fig. 3.1).
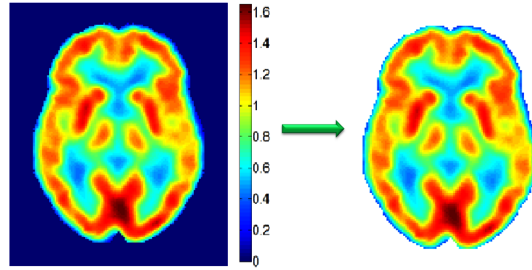

Figure 3.1: Illustrative example of a brain FDG-PET scan (transaxial view). On the left image, the FDG-PET scan with a colorbar indicates the intensity of each voxel, i.e., the FDG uptake. Blue colors represent lower levels of CMRgl whereas warm colors indicate higher levels of CMRgl. On the right, an image of the same FDG-PET scan with the extracerebral voxels removed is illustrated.

Although the whole brain approach has been reported to be more beneficial for most advanced stages of the disease [45], the ROI approach has been widely used due to its advantages. Extracting only the most discriminative brain regions leads to a dimensionality reduction of the features vectors, which have a great influence in the algorithm performance and can save computational cost. On the other hand, this method requires the choice and extraction of the regions, which is a time-consuming and user-dependent task. In order to overcome these drawbacks, several anatomical atlas based on MRI scans have been developed (for example, [54], [55]) as well as automatic segmentations of certain regions of the brain, such as the hippocampus (for example, [56], [57]). In the present thesis, an atlas developed by Desikan et al. [55] was used in the multi-region approach. This atlas automatically segmented the whole brain into 34 cortical ROIs in each hemisphere plus the brainstem, using a dataset of 40 MRI scans (Fig. 3.2). A total of 69 regions across whole brain are encoded in the atlas. For the regional analysis, the mean of the voxels intensity of each region was extracted, a totaling of 69 features for each subject. In Fig. 3.3, an example of the brain segmentation of an FDG-PET scan using the adopted atlas is illustrated.

3.2.2 Longitudinal Features

The incorporation of longitudinal information for the automatic diagnosis of AD only recently begun to receive some attention (see Chapter 2), therefore feature extraction for longitudinal data is not a widely investigated topic. In the present thesis, a simple and intuitive method for longitudinal feature extraction is proposed. The longitudinal information is comprised by the baseline, 6-month, 12-month and 24-month FDG-PET follow-up scans of each subject. Similarly to the extraction of the cross-sectional features, the extraction of the longitudinal features was performed using a voxel-based and regional-based analysis. In the voxel-based approach, the differences between baseline VI and VI
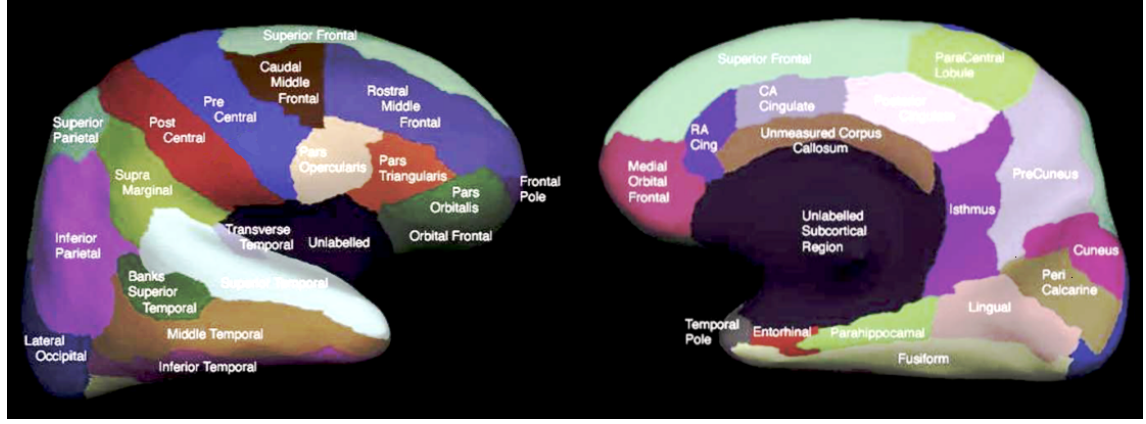
Figure 3.2: Representation of some of the labeled ROIs in the brain hemisphere encoded in the Atlas developed by Desikan et al. [55]. The left image illustrates the lateral view of the hemisphere and the right image illustrates the medial view of the hemisphere. Adapted from: Desikan et al. [55].

of corresponding voxels in the follow-up scans were extracted. Hence, different sets of features were obtained: VI differences between baseline and 6-month scans, VI differences between baseline and 12-month scans and VI differences between baseline and 24-month scans. The result of this procedure, using the 24-month follow-up scans, is illustrated in Fig. 3.4. As one can observe, the brain extension and value of the differences over a 24-month period are larger for AD patients compared to the CN subjects. Furthermore, a closer look can detect a decline in the intensity of the images between groups at the baseline and 24-month follow-up scans. Regarding the regional-based analysis, all time points scans were first segmented into ROIs, using the above mentioned anatomical atlas. Then, the mean VI of each region, in each time-point scan was computed. Finally, the regional differences between the baseline and the follow-up scans were extracted for each subject, originating different sets of regional longitudinal features.
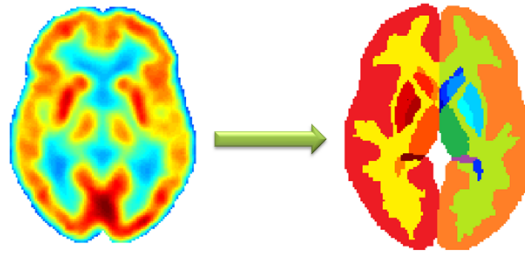
Figure 3.3: A transaxial view of a FDG-PET scan illustrating the brain segmentation into different ROIs. On the left, an FDG-PET scan containing only the intracerebral voxels is present. On the right, the anatomical masked segmentation into different cortical regions with a distinction between the regions localized in the left and right hemispheres is illustrated. Different colors correspond to different regions.

It is assumed that the differences extracted reflect the CMRgl decline along time. Hence, higher differences correspond to brain regions where neuronal degeneration is more intense, and lower differences correspond to brain regions that are relatively spread by the disease process along time. It is also expected that the signal intensity differences between the baseline and 24-month scans are more pronounced than the signal intensity differences between baseline and 6-month or 12-month scans, since AD is a progressive disease and neuronal degeneration worsens over time.
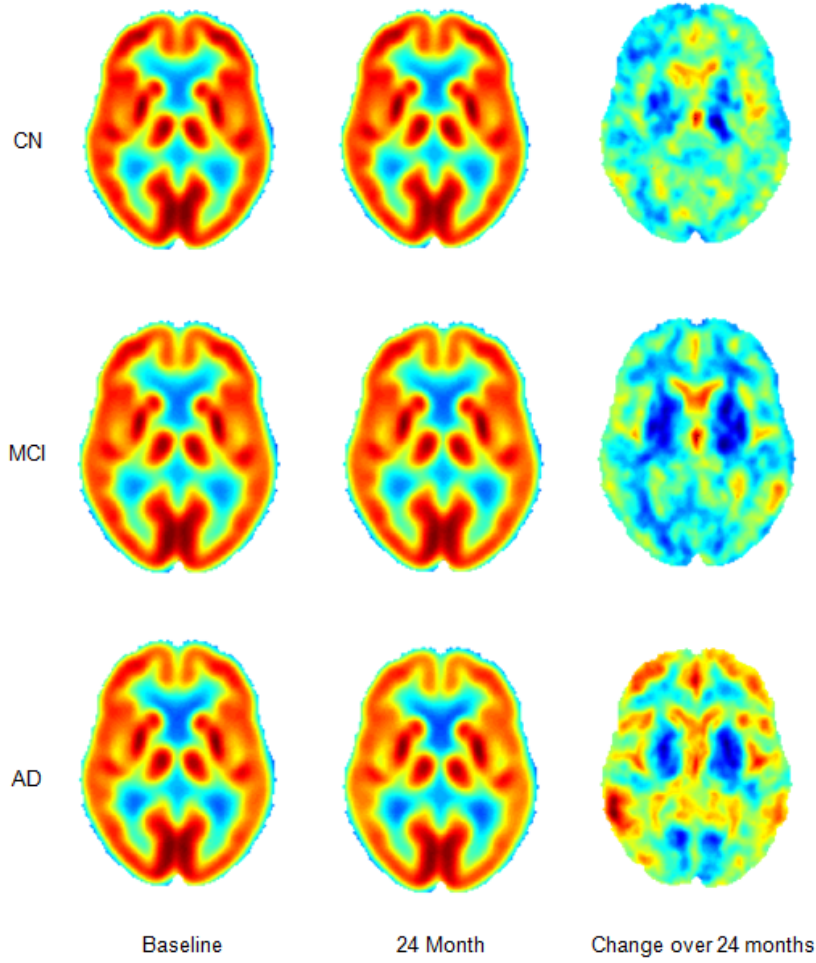
Figure 3.4: Transaxial sections of FDG-PET brain scans at the baseline, 24-month and the VI change over 24 months from CN, MCI and AD subjects. The 24-month changes are computed by subtracting the VI of the 24-month scans from the VI of the baseline scans. For the cross-sectional scans, warmer colors indicate higher FDG uptake and blue colors indicate lower FDG uptake. In the third column, the warmer colors in the images indicate higher metabolic differences along the 24-month period and blue colors indicate relatively preservation of the metabolism along this period.

3.3   Feature Selection

Feature selection is a crucial data preprocessing step, in particular when the number of features highly exceeds the number of available samples, a situation commonly refered to as the *curse of dimensionality*. Under these conditions, the classifier tends to overfit the data and the generalization ability is compromised, leading to a deterioration of the classifier performance. Furthermore, the features extracted may contain noisy data, irrelevant and redundant information, which can have a great influence on the final outcomes. Additionally, a high-dimensional input features tends to be computationally expensive. Hence, the ultimate goal of the feature selection step is to select the smallest subset of features that maximally increases the performance of the classifier, providing an acceptable trade-off between the results and the computational cost. Besides improving the accuracy results, feature selection techniques also provides a better insight of the processes described by the data, since only the relevant features are used as input for the classifier.

Feature selection techniques for machine learning typically fall into two categories, namely filter methods and wrapper methods, depending on how the algorithm interacts with the classifier [58].

Filter methods are directly applied on the dataset selecting a specific number of features according to some ranking score. Wrapper methods assess the quality of a set of features guided by the outcome of the classifier itself. Wrapper methods usually yield higher predictive accuracies than filters since the former optimize the feature selection process according to the results obtained with the learning algorithm employed. However, applying a learning algorithm to evaluate the classification results using every set of features is extremely expensive from a computational point of view. On the other hand, filter methods that treat each feature independently scoring them according to its discriminative power, are very efficient and fast to compute. However, a feature that is not useful alone, may provide valuable information when combining with other features, and unlike wrappers, the filter methods do not capture this relationship. Still, the advantages of filter methods, especially in a voxel-based analysis where the number of features are extremely high, may outweigh their disadvantages in some particular situations. For these reasons, only filter methods are explored in the present thesis and they are described in the following subsections.

3.3.1 Correlation Coefficient

The feature selection scheme based on the correlation coefficient consists on computing the correlation between each feature and the class label. The value obtained reflects the relevance of each feature to identify the clinical group.

In the present work, the correlation coefficient used was the Pearson's linear correlation. Consider a vector $\mathbf{v}_j$ containing the intensities of the j-th voxel across K subjects and a vector $\mathbf{y}$ containing the correspondent class label for each subject ($\mathbf{y}$ = -1 for healthy control subjects and $\mathbf{y}$ = 1 for pathological subjects (AD or MCI)). Pearson's correlation coefficient r at the j-th voxel is computed as follows:

$$r(\mathbf{v}_j, \mathbf{y}) = \frac{\sum\limits_{i=1}^{K}(\mathbf{v}_{ij} - \overline{\mathbf{v}}_j)(\mathbf{y}_i - \overline{\mathbf{y}})}{\sqrt{\sum\limits_{i=1}^{K}(\mathbf{v}_{ij} - \overline{\mathbf{v}}_j)^2}\sqrt{\sum\limits_{i=1}^{K}(\mathbf{y}_i - \overline{\mathbf{y}})^2}}, \tag{3.1}$$

where $\mathbf{v}_{ij}$ is the value of the j-th voxel in the i-th FDG-PET image, and $\mathbf{y}_i$ is the class label (-1 or 1) of the corresponding i-th image. Hence, each voxel is characterized by a correlation coefficient with values comprised between -1 and 1, where -1 corresponds to a total negative correlation, 1 indicates a total positive correlation and 0 means no relationship between the feature and the label. Features with highest absolute correlation coefficient are selected as input to the classifier.

3.3.2 t-Test

The t-test is a statistical test used to assess whether the means of two populations are statistically different from each other. In other words, this method tests the null hypothesis that data from two populations comes from independent samples from normal distributions with equal means and equal but unknown variances. The alternative hypothesis is that the two populations have different means. The t-test returns a t-value which computes the mean differences between the two populations. Thus, the higher the value of t, the larger are the confidence to reject the null hypothesis. This test also

returns a p-value which is a scalar with values comprised between 0 and 1, inclusive. The p-value reflects the probability, under the assumption that the null hypothesis is true, of obtaining a given result, or one more extreme, by chance. The null hypothesis is rejected when the p-value is lower than an established significance level, usually 0.05.

The t-test has been widely used as feature selection method in order to select the most discriminative voxels between CN and AD or MCI subjects, in several published papers (for example [25], [41], [59]). This statistical test compares images from different clinical groups using a voxel-by-voxel analysis. A SPM of the group's differences is therefore generated, identifying regions where the pathological group has reduced CMRgl as compared to the cognitively healthy group. In the present thesis, an unpaired two-sample t-test was conducted between different pairs of clinical groups (CN vs AD, CN vs MCI, MCI vs AD). The t-value is estimated as follows:

$$t_j = \frac{\overline{I}_{y_{-1}} - \overline{I}_{y_1}}{S_{y_{-1}y_1}\sqrt{\frac{1}{k_{-1}} + \frac{1}{k_1}}}, \tag{3.2}$$

where t_j is the t-value of the j-th voxel, $\overline{I}_{y_{-1}}$ and $\overline{I}_{y1}$ denotes the mean voxel intensities of the FDG-PET images of the subjects labeled as $y = $ -1 and $y = 1$, respectively, k_{-1} and k_1 denotes the number of subjects labeled as $y = $ -1 and $y = 1$, respectively. $S_{y_{-1}y_1}$ is the estimation of the common standard deviation of the two samples and it is calculated as:

$$S_{y_{-1}y_1} = \sqrt{\frac{(k_{-1})S_{y_{-1}}^2 + (k_1 - 1)S_{y_1}^2}{k_{-1} + k_1 - 2}}, \tag{3.3}$$

where $S_{y_{-1}}$ and S_{y_1} are the sample standard deviation image for the population labeled as y=-1 and y=1 respectively, and it is defined as (i=-1,1):

$$S_i = \sqrt{\frac{1}{k_i - 1}\sum_{j=1}^{k_i}(I_j - \overline{I}_{y_i})^2} \tag{3.4}$$

In the present work, the voxels of the FDG-PET images are ranked using the absolute t-value obtained by the t-test. A higher t-value indicates significant differences between the mean of the healthy control group and the pathological group (MCI or AD) which is related to the decline of CMRgl. Hence, the t-test selects the most discriminative features between clinical groups, i.e., the voxels were the brain FDG uptake is compromised due to the disease process.

3.3.3 Mutual Information

Mutual information is a nonparametric measure of the mutual dependence of two variables, i.e., it represents the reduction in uncertainty about a certain variable x when the value of a variable y is known. The mutual information as feature selection method has been proposed [60], since it assesses the information content of the features.

The mutual information of two random variables x and y is defined in terms of their respective

probabilistic density function $p(x)$ and $p(y)$ and their joint density function $p(x,y)$:

$$\text{MI}(x,y) = \int_x \int_y p(x,y) \log \frac{p(x,y)}{p(x),p(y)} dx dy, \tag{3.5}$$

Hence, $\text{MI}(x,y)$ estimates the dependency between the density of variable x and y. Observing Eq. 3.5, if x and y are independent, then $p(x,y) = p(x)p(y)$, therefore their mutual information is 0, which means they contain no information about each other.

When the variables are discrete, their probabilities are estimated by frequency counts and Eq. 3.5 is rewritten as:

$$\text{MI}(x,y) = \sum_x \sum_y P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \tag{3.6}$$

The probabilities densities can be approximated using histograms. In the present thesis, the features with higher mutual information values were selected, since they reflect a stronger dependency on the class label, hence a higher discriminative power.

3.4 Classification - Support Vector Machines

Support Vector Machines (SVMs) are a set of machine learning algorithms introduced by Boser, Guyon and Vapnik in 1992 [61]. Currently, SVM is one of the most used classifier in a wide range of applications, due to its high performance even when dealing with high-dimensional data. SVMs algorithms are based on supervised learning that uses training data to build a model able to predict the class of a new sample. The evaluation of the classifier performance along with parameter optimization is essential in order to validate the obtained model. Nested cross-validation is one of the most popular methods and the one used in the present work. A detailed description of this procedure and the basic concepts and mathematics of SVMs are now presented.

3.4.1 Basic Concepts and Mathematics

The concept behind SVM algorithm relies on the construction of a hyperplane which separates a set of binary labeled training data with maximal margin between the vectors of the two classes (known as maximum margin hyperplane). Only a small number of training vectors are necessary to determine this margin in order to construct the hyperplane, the so called support vectors. The construction of this hyperplane is illustrated in Fig. 3.5, for two linearly separable classes which is the simpler possible scenario.

If the training data are separated by an optimal hyperplane without errors, the expected probability of committing an error on a test example is given by the ratio between the value of the expected number of support vectors and the number of training vectors:

$$\mathbf{E}[P(error)] \leq \frac{\mathbf{E}\ [number\ of\ support\ vectors]}{number\ of\ training\ vectors} \tag{3.7}$$

As one can observe, the expected error does not explicitly depend on the dimensionality of the feature vector. Hence, if a small number of support vectors relative to the training data size is used to construct
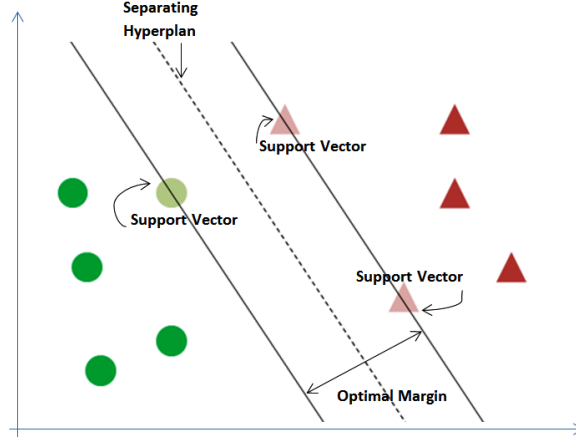
Figure 3.5: Illustration of the concept of maximum margin hyperplane used by SVM algorithm. The algorithm searches for a boundary that maximizes the distance of the closest points of each class. These points are called support vectors.

the hyperplane, the generalization ability of the classifier is large, even for high-dimensional data. The mathematics behind the implementation of SVM algorithm will be addressed next. A more detailed explanation can be found in [62].

Consider a set of linearly separable training samples $S = \{(x_1, y_1), ..., (x_k, y_k)\}$, where $x \in \mathbb{R}^n$ denotes the input space and $y \in \{-1, 1\}$ the output domain for binary classification. The optimal hyperplane in the feature space that separates the two classes can be parameterized by its normal vector $\mathbf{w}$ and a constant b:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{3.8}$$

Given this hyperplane, the decision function is defined as:

$$f(x) = sign(\langle \mathbf{w} \cdot \mathbf{x} \rangle + b) \tag{3.9}$$

The two classes are linearly separable if there is a vector $\mathbf{w}$ and a scalar b that satisfies the following conditions:

$$\mathbf{w} \cdot \mathbf{x_k} + b \geqslant 1 \text{ if } y_k = 1,$$
$$\mathbf{w} \cdot \mathbf{x_k} + b \leqslant -1 \text{ if } y_k = -1, \tag{3.10}$$

The inequalities 3.10 can be rewritten in the form:

$$y_i(\mathbf{w} \cdot \mathbf{x_k} + b) \geqslant 1 \ \forall k \tag{3.11}$$

Since the optimal hyperplane is the one that separates the two classes with a maximal margin, this distance is given by:

$$d((w, b), x_k) = \frac{y_k(\mathbf{w} \cdot \mathbf{x_k} + b)}{|| \mathbf{w} ||} \tag{3.12}$$

In order to obtain the minimum distance between support vectors, which corresponds to maximizing the margin, Eq. 3.12 must be minimized by optimizing the parameters $\mathbf{w}$ and b. This can be formulated by the following optimization problem:

$$\min_{\mathbf{w},b} \quad \frac{1}{2}\mathbf{w^T w}$$

$$\text{subject to} \quad y_k(\mathbf{w} \cdot \mathbf{x_k} + b) \geqslant 1, \ \forall k \tag{3.13}$$

Eq. 3.13 is an example of a quadratic programming problem where the aim is to minimize a quadratic function subject to a set of linear inequality constrains. In order to solve this constrained optimization problem, Langrange multipliers are introduced:

$$L(\mathbf{w}, b, \Lambda) = \frac{1}{2}\mathbf{w^T w} - \sum_{k=1}^{K} \alpha_k [y_k(\mathbf{w^T x_k} + b) - 1] \tag{3.14}$$

where $\Lambda = (\alpha_1,...,\alpha_k)$ is the vector of non-negative Lagrangian multipliers correspondent to the constrains in Eq. 3.13. The aim is to find the $\mathbf{w}$ and b values which minimize and the α which maximizes Eq. 3.14. This can be achieved by differentiating L with respect to $\mathbf{w}$ and b and setting the derivatives to 0:

$$\frac{\partial L(\mathbf{w}, b, \Lambda)}{\partial \mathbf{w}} = 0 \implies \mathbf{w} = \sum_{k=1}^{K} \alpha_k y_k x_k \tag{3.15}$$

$$\frac{\partial L(\mathbf{w}, b, \Lambda)}{\partial b} = 0 \implies \sum_{k=1}^{K} \alpha_k y_k = 0 \tag{3.16}$$

Substituting 3.15 and 3.16 back in 3.14, the dual representation of the problem is obtained:

$$W(\Lambda) = \sum_{k=1}^{K} \alpha_k - \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{K} y_k y_j \alpha_k \alpha_j (x_k^T x_j), \tag{3.17}$$

where the vectors appears only inside of dot products. In order to obtained the solution, Eq. 3.17 is maximized with regard to α. Finally, the Karush-Kuhn-Trucke condition has to be satisfied according to optimization theory, which states that $\alpha_k = 0$ is verified for points inside the class and do not contribute to the solution. On other hand, $\alpha_n \neq 0$ is verified for points that lie on the margin (support vectors) and do contribute to the solution.

When no linear separation of a dataset is possible, SVM are combined with kernel techniques. The data is mapped into a higher dimensional space becoming linearly separable and the hyperplane corresponds to a non-linear decision boundary in the input space. The dual representation of linear relations previously described, allows the training algorithm to depend on the data only through inner products between all pairs of observed point without explicitly defining its coordinates. Therefore, if the kernel function is defined by $\mathrm{K}(x_k, x_j) = \Phi(x_k) \cdot \Phi(x_j)$ then there is no need to know the mapping

Φ explicitly. The most commonly used kernel is the Gaussian Radial Basis Function (RBF):

$$K(x_k, x_j) = e^{(-\gamma \|x_k - x_j\|^2)}, \tag{3.18}$$

where $\gamma > 0$ defines the width. There are several other kernels, and the choice of the kernel to use depends on the characteristics of the classification problem, such as the number of features and samples available. For a high dimensional problem, the linear kernel has been proved to yield better classification results, although in some cases the RBF kernel is more suitable [63].

All formulations presented so far were made based on the assumption that a linear separation of the data is possible either on the input space or in the high-dimensional space. For non-separable classes, the optimization process is modified so that misclassification of the features vectors are allowed. Thus, slack variables ε are introduced to measure the degree of misclassification and now the optimization problem involves maximizing the margin and minimizing the error:

$$\min_{\mathbf{w}, \varepsilon, b} \left\{ \frac{1}{2} \mathbf{w^T w} + C \sum_{k=1}^{K} \varepsilon_k \right\},$$
$$\text{subject to} \quad y_k(\mathbf{w} \cdot \mathbf{x_k} + b) \geqslant 1 - \varepsilon_k$$
$$\text{and } \varepsilon_k \geqslant 0, \ \forall k \tag{3.19}$$

Where the parameter C is a positive constant that controls the trade-off between allowing training errors and forcing rigid margins. Hence, a soft margin is created which permits some misclassifications. Increasing the value of C increases the penalty for misclassifying points, forcing the creation of a more accurate, but less general model. In the present thesis, the SVM classifier was applied using LIBSVM, a toolbox developed by Chang and Lin [64].

3.4.2 Classifier Performance - Nested Cross-Validation

The performance assessment of the classifier is performed in order to optimize the parameters and to evaluate the validity of the obtained model. When a large dataset is available, the data is split into 3 disjoint groups, namely, the training set, the validation set and the test set. The training set is used to train the algorithm and its parameters are chosen according to the best results obtained using the validation set. Then the classifier predicts the class of the test data. Thus, the performance of the classifier is estimated by computing the mean accuracy obtained in a sufficiently large test dataset. It is of utmost importance to ensure that the data used to train the algorithm is not the same data used to evaluate its performance, since it would yield to an overoptimistic estimative of the error. However, when the size of the dataset is small, i.e., the number of features broadly exceeds the number of samples, it is not recommended to leave out data from the learning stage since sufficient samples are needed to construct the classification model. Cross-validation strategy was designed to overcome the lack of a sufficiently large dataset. The most commonly used method is the *k-fold* cross-validation, where data is randomly partitioned into k subsets. One of the sets is used as validation set and the remaining k-1 subsets are used for training. This procedure is repeated for different left out portions. Then, the average error of all iterations is computed and it can be interpreted as the true error of the classifier.

In 2006, Varma and Simon [65] demonstrated that using the cross-validation strategy to estimate the error of a classifier that has itself been tuned using the cross-validation leads to a significantly biased estimation of the true error. Hence, a nested cross-validation method, schematized in Fig. 3.6, was proposed.
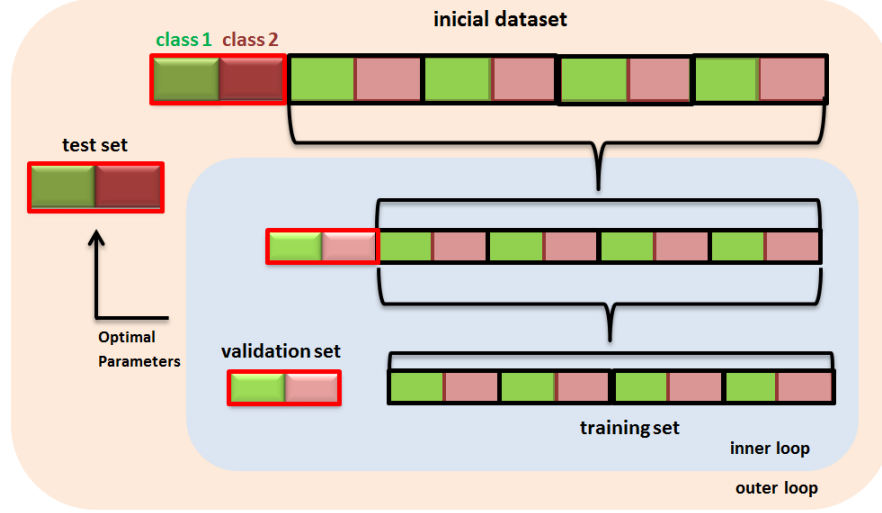
Figure 3.6: Illustration of the nested cross-validation scheme. This procedure consists in an inner and an outer loop. Two different classes are represented by green (class 1) and red (class 2) colors. In the outer loop, the initial dataset is split into k folds (in the example there are 5 folds). In the present schematic representation, the number of samples of each class is the same in each fold. One of the folds is chosen as test set and it is left out of the inner loop. The remaining data is again divided into several (k or other number) folds, and one of the folds is chosen as validation set and the remaining k-1 folds constitute the training data. The tunning of the parameters is achieved by performing a standard cross-validation procedure. The averaged accuracy results are computed leading to the identification of the optimal parameters. The model with these parameters is then used to the classification of the test set, with a model learned from all the data entered in the inner cycle. This process is repeated k times, each time leaving a different fold out of the inner cycle.

This method consists of an inner and an outer loops. Firstly, in the outer loop, the initial dataset is partitioned into k folds, usually 5 or 10 folds are used. The partition may be random or take into account the number of samples of each class in each fold. One of the folds is left out of the inner cycle and used as test set and the remaining k-1 folds are used as training data entering the inner loop. The training data is again divided into k or other number of folds. One of the folds is used as validation set and the remaining folds are used as training data. A standard cross-validation procedure is performed within this inner loop, and repeated k-times, where in each iteration, different folds are chosen as validation set. The optimal parameters are chosen according to the mean of the classification accuracy computed across all iterations. Finally, the model obtained is used to classify the test set. Since the test set did not enter the inner loop where the parameter optimization took place, this procedures leads to an unbiased estimation of the classifier performance. In the present thesis, a nested 10-fold cross-validation strategy was applied.

4

Experimental Results

Contents

The present chapter begins by describing the characteristics of the study population and the neuroimaging data analyzed. The imaging pre-processing conducted by ADNI is also described in Section 4.1. The experimental design is explained in Section 4.2 and the results obtained are presented and discussed in Section 4.3.

4.1 Neuroimaging Data

The neuroimaging data used in the present work were retrieved from the ADNI database. The primary goal of ADNI is to develop imaging, clinical, genetic and biochemical biomarkers for baseline characterization and longitudinal analysis of AD. The study began in 2004 and it enrolled, in its first phase (ADNI1), 400 MCI patients, 200 individuals suffering from AD and 200 elderly control subjects. About half of these subjects underwent FDG-PET scans at the baseline and, according to the ADNI protocol, AD patients were also submitted to follow-up scans at month 6, 12 and 24, CN subjects at 6, 12, 24 and 36 and MCI patients at 6, 12, 18, 24 and 36 month. More detailed information on the ADNI study can be found at http://adni.loni.usc.edu/.

4.1.1 Participants

Since the ADNI protocol did not include scans at month 18 and 36 for all the clinical groups, only FDG-PET scans at baseline, month 6, 12 and 24 were retrieved for the present study. Table 4.1 summarizes the demographic (number of subjects, mean age and percentage of males) and clinical information (mean MMSE score at baseline and follow-ups, and the distribution of the CDR scores at baseline) of the study population. Participants enrollment was also conditioned by the CDR score: 0 for CN, 0.5 for MCI and 0.5 or more for AD patients. To ensure that the classification was only based on disease-specific imaging information rather than the intrinsic age and gender captured by the scans, a 2 sample t-test was performed between different clinical groups for age and gender. All the results (CN vs AD, CN vs MCI, MCI vs AD) of the t-test, for the null hypothesis that the data have equal means and variances, both for age and gender, yield p-values higher than 0.05, which means that the null hypothesis can not be rejected. Hence, one can conclude that the differences of age and gender between clinical groups are not statistically significant.

Group	AD	MCI	CN
Number of subjects	48	109	66
Age ($\mu \pm \sigma$)	76.5 ± 6.7	74.8 ± 7.1	75.9 ± 4.5
Sex (% of Males)	56.3	63.3	62.1
MMSE at baseline ($\mu \pm \sigma$)	23.4 ± 2.0	27.2 ± 1.6	29.1 ± 1.0
MMSE at month 6 ($\mu \pm \sigma$)	22.5 ± 3.2	26.9 ± 2.5	29.1 ± 0.8
MMSE at month 12 ($\mu \pm \sigma$)	21.3 ± 3.9	26.6 ± 2.7	29.1 ± 1.3
MMSE at month 24 ($\mu \pm \sigma$)	20.0 ± 4.9	25.7 ± 3.5	29.0 ± 1.1
CDR = 0 (%)	0	0	100
CDR = 0.5 (%)	35	100	0
CDR = 1 (%)	65	0	0

Table 4.1: Demographic and clinical information of the study population. μ and σ stand for mean and standard deviation. MMSE - Mini-Mental State Examination; CDR - Clinical Dementia Rating.

4.1.2   Imaging Pre-Processing

The FDG-PET scans of the ADNI database were acquired according to one of three standardized protocols [66]: 30-minute static (a single 30-minute frame, 30 to 60 minutes post-injection), 30-minute dynamic (six 5-minute frames, 30 to 60 minutes after FDG injection) and 60-minute dynamic (60 minutes dynamic protocol composed of 33 frames starting at the moment of the tracer injection). In order to standardize the PET data acquired with different systems, the ADNI researchers performed a series of preprocessing steps including [67]: registration of the several scans, acquired during a single visit, to each other and averaging; reorientation of the averaged image to a common spatial orientation and resample using a grid having 1.5 mm cubic voxels; filtration of the reoriented and resampled image with a scanner-specific function to provide images with an apparent resolution similar to the lowest resolution scanners used in ADNI.

The ADNI preprocessing procedure also includes the intensity normalization of the FDG-PET images. Therefore, in order to investigate the intensity normalization methods proposed in the present work, it was necessary to retrieve the non-normalized FDG-PET images from de ADNI database. However, these images were not aligned, so in order to make voxel-wise comparisons, all images had to be warped into the MNI152 standard space. Hence, the correspondent 1.5T MRI scans of each subject at each time-point were also retrieved from ADNI. Similarly to the FDG-PET images, the MR images had undergone a series of preprocessing steps by ADNI researchers in order to eliminate artifacts. These included corrections for gradient non-linearity distortions using a scanner-specific function and corrections for non-uniformities in the image intensity.

The procedure to warp all the images into the MNI152 standard space was achieved by performing the following steps. Firstly, the brain tissue in all MR images was extracted (skull-stripping) and segmented into white-matter and gray-matter. The extraction of brain tissue was performed with FreeSurfer [68] and the tissue classification was conducted with SPM8 [69]. Secondly, all PET images were co-registered with the corresponding skull-stripped MR images using SPM8. In order to conduct these co-registrations, rigid-body transformations (6 degrees of freedom) and an objective function based on the normalized mutual information between the two images were applied [70]. Thirdly, the MR images acquired for each subject at baseline and follow-up scans were non-linearly registered into a subject specific template using the DARTEL toolbox from SPM8 [71]. Finally, all the baseline MR images were non-linearly registered to an inter-subject template using DARTEL, and the resulting template was mapped to the MNI-ICBM 152 non-linear symmetric atlas (version 2009a) [72] using an affine transformation. After completing the aforementioned steps, the original PET and MR images were resampled into the MNI152 standard space with a 1.5×1.5×1.5 mm resolution using the appropriate composition of transformations. The final images were represented by a 121×145×121 matrix.

4.2   Experimental Design

The main goal of the present work was to evaluate the value of longitudinal information for the classification of CN, MCI and AD subjects. Additionally, the influence of the intensity normalization of the FDG-PET images on the performance of the classification task was investigated. The extraction

of features was performed using a voxel-wise and a multi-region analysis and different filter methods were applied in order to select the most relevant features. The methods proposed were evaluated in both cross-sectional and longitudinal data and, in order to assess their performance, classification experiments were conducted for CN vs AD, CN vs MCI and MCI vs AD. A detailed description of the experimental design and the implemented methods is now presented.

Intensity Normalization Methods

Three different methods for the intensity normalization of the FDG-PET images were tested, namely, the CGM, the regional mean and the reference cluster approaches. For the regional mean intensity normalization, the cerebellum and the brainstem were used (see Chapter 3). Regarding the reference cluster intensity normalization, the implemented method was based on the approach proposed by Yakushev et al., [34] (see Chapter 3), but introducing a different process for the selection of the reference cluster. The original method performs a search within the clusters obtained, and the cluster containing the voxel with the highest t-value is used as reference cluster. In this work, the mean of the t-values of each selected cluster is extracted, and the cluster with the highest t-value mean is used as reference. Hence, a more robust method is introduced, since the choice of the reference cluster is based on the mean of all the t-values of a certain region rather than just a single t-value. Furthermore, since the size of the images used in the work of Yakushev et al. [34] ($91 \times 109 \times 91$) is different from the images used in present work ($121 \times 145 \times 121$), the values of the parameters t and r had to be adjusted. A large value of r implies that larger spatial extent of contiguous voxels is required, and a large value of t means than only regions with large intensity differences between two groups are selected. Consequently, the number and area of the clusters selected will be small. On the other hand, if both parameter values are small, the area of the selected clusters will be large and the reference cluster for normalization will comprise a larger region. Therefore, it is necessary to find the most suitable combination of r and t values. Due to the extreme computational cost implied in the optimization of the parameters of the reference cluster method in each classification task, a grid search was conducted within the training set in order to estimate de best parameters for each pair of clinical groups, at each time-point data (Table 4.2).

	CN vs AD			CN vs MCI			MCI vs AD		
	t	r	Ref. Cluster	t	r	Ref. Cluster	t	r	Ref. Cluster
Baseline	2	343	SMC	1	27	Cerebellum	3	125	SMC
Month 6	3	343	Cerebellum	1	27	Cerebellum	3	125	SMC
Month 12	4	729	SMC	1	27	SMC	3	343	SMC
Month 24	4	1331	SMC	2	27	Brainstem	3	343	Brainstem

Table 4.2: Optimal parameters (t and r) for the reference cluster normalization method and the localization of the reference cluster. The optimal values were achieved by performing a grid search over t (range: 1 to 5) and r (range: 3^3 to 11^3) using a 10-fold nested cross-validation and performing 10 iterations. The (t,r) pair with the highest mean cross-validation accuracy was selected.

As one can observe in Table 4.2, the t and r values tend to increase over time, indicating that the hypermetabolic regions due to the CGM intensity normalization procedure in pathological subjects compared to controls, increases in terms of intensity and spatial extent. This can be explained by the fact that the CGM between the different clinical groups tends to diverge along time. Furthermore,

the parameter values are generally higher for CN vs AD compared to other two pairs, reflecting the larger CGM differences between these subjects. Regarding the localization of the reference cluster, it is situated mainly in brain areas referred in literature as conserved regions (see Chapter 3): the SMC, the cerebellum and the brainstem. The proposed intensity normalization methods were tested using the best $F \in [1000:1000:15000]$ features of the baseline images selected with the t-test.

Feature Extraction

The VI of the FDG-PET images were extracted by using a voxel-based and a multi-region approaches, for both cross-sectional and longitudinal data. In the voxel-wise analysis, the voxel intensities of the baseline and follow-up images were obtained and, since the FDG-PET images were represented by a $121 \times 145 \times 121$ matrix, a total of 2122945 voxels were available. A mask was applied in order to exclude all the extracerebral voxels, which contains no information, and a total of 557780 intracerabral voxels were obtained. Regarding the multi-region analysis, the mean of the voxel intensities was extracted for each of the 69 regions, at baseline and follow-up scans. For the longitudinal analysis, the VI and regional differences of each follow-up scans relative to baseline were computed. Therefore, longitudinal features comprised three different sets, namely: the differences between the baseline and the 6-month scans, the baseline and 12-month scans and the baseline and the 24-month scans.

Feature Selection Methods

The t-test, linear correlation and mutual information were used to select the most discriminative cross-sectional and longitudinal features. The performance of the feature selection methods were assessed using the baseline scans and the differences over a 24-month period. The features were sorted in descending order according to the value obtained in each method (the t-value, the correlation coeficient and the mutual information). In the voxel-based analysis, the best $F \in \{100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 30000\}$ features were used as input to the classifier. Although the features generated by the multi-region approach did not suffer from the *curse of dimensionality*, some of the features might contain no relevant information. Hence, the feature selection methods were also applied to select the best regional features. The 69 regional features were ranked according to each feature selection method, and the best $F \in [1:1:69]$ features were used for classification.

Cross-sectional and Longitudinal Imaging Data

Firstly, classification experiments were performed using the 6-month, 12-month and 24-month imaging data and using the VI and regional changes over 6 months, over 12 months and over 24 months. For the multi-region analysis, all 69 regions were used. For the voxel-based analysis, 1000 features were selected with t-test in the cross-sectional data. For the longitudinal data, the best 1000 features were selected in the follow-up scans, and the VI change of the selected voxels were used. Then, the cross-sectional imaging data were combined with the longitudinal data by simple concatenation of features. Each single time-point imaging data was combined with the changes between the baseline and the respective follow-up imaging data. Additionally, the baseline data were also combined with all longitudinal sets. The features were selected in the cross-sectional data with the t-test, and an optimization of the number of features, with $F \in [1000:1000:15000]$ within the training set was performed. For the

longitudinal data, the best 15000 features were firstly selected with the t-test in the follow-up data. Then, assuming that voxels with higher differences along the follow-up period correspond to brain regions with higher atrophy rate due to the progress of the disease, the longitudinal VI differences of all 15000 features were sorted in descending order. Finally, an optimization of the number of features with $F \in [1000{:}1000{:}15000]$ was also performed.

Classification and performance evaluation

In order to evaluate the performance of the methods proposed, the classification was performed with the SVM classifier for CN vs AD, CN vs MCI and MCI vs AD. For the voxel-based analysis, the number of features was much larger than the number of instances, thus, a linear kernel was chosen since mapping the input data into a high-dimensional space using more complex kernels is not necessary. In order to define the range of the parameter C, the classification was first assessed using a wide range of values. Fig. 4.1 shows the accuracy obtained when using different values of C parameter, ranging from 2^{-20} to 2^{20} for CN vs AD, CN vs MCI and MCI vs AD classification. These results were obtained using the best 1000 VI features selected with t-test in the baseline data with a 10-fold cross-validation strategy, repeated 10 times. As one can observe in Fig. 4.1, the best accuracy was achieved for C values between 2 and 2^{14} for all three classification tasks. Hence, the C parameter was set to range between 2 and 2^{14} and was tunned in each classification task using a 10-fold nested cross-validation strategy within the training set.
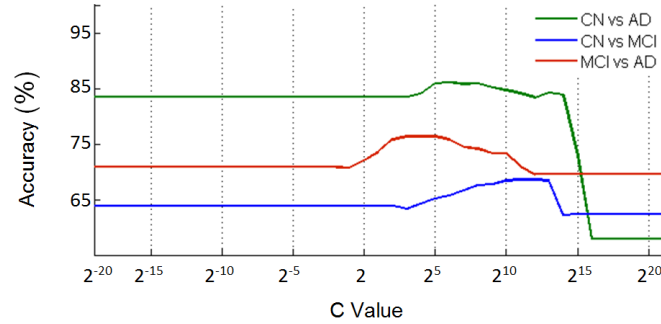
Figure 4.1: Classification accuracy varying the C parameter of the SVM algorithm for CN vs AD, CN vs MCI and MCI vs AD classifications. The results were obtained using 1000 VI features selected with the t-test in the baseline data and adopting a 10-fold nested cross-validation strategy.

For the multi-region analysis, since the number of features is relatively well balanced with the number of subjects, a RBF kernel was chosen along with the optimization of the C (range: 2 to 2^{14}) and γ (range: 10^{-5} to 5^{-2}) parameters by performing a grid search adopting a 10-fold nested cross-validation strategy within the training set. In order to define the range of γ values, a similar procedure to the one used with the C parameter, was conducted. For all classification tasks and each set of features, a 10-fold cross-validation strategy was used to assess the classification accuracy, i.e., the dataset was split, 90% was used for training and 10% was used for testing. This process was repeated 10 times in order to avoid any bias due to the random partition of the dataset in cross-validation. The comparison of performance of all proposed classifiers will be based on the average accuracy and standard deviations obtained. In addition, a 2-sample t-test using a significance level of 0.05, was also performed to assess the statistical significance of the results obtained with the combination of cross-sectional data and longitudinal data compared to use the cross-sectional data alone.

4.3 Results

The results are organized with the following subsections: the intensity normalization methods (Subsection 4.3.1), features selection methods for both cross-sectional and longitudinal data (Subsection 4.3.2) and the classification results using the cross-sectional combined with the longitudinal data (Subsection 4.3.3).

4.3.1 FDG-PET Intensity Normalization Methods

The reference cluster obtained between CN, MCI and AD groups, using the parameters displayed in Table 4.2 for the baseline data, is presented in Fig. 4.2. Note that the reference clusters were estimated using the whole sample of subjects for illustration purposes only (in the classification process the reference cluster was always obtained within the training set).

As one can observe in Fig. 4.2 (a), the reference cluster (areas marked in red) obtained between CN and AD subjects, is mainly situated in the SMC, although other clusters (areas marked in black) are also identified in the brainstem and cerebellum. For the CN vs MCI, the reference cluster is in the cerebellum and brainstem regions, and for MCI vs AD, a small reference cluster is detected in the SMC.

(a) CN vs AD   **(b)** CN vs MCI   **(c)** MCI vs AD
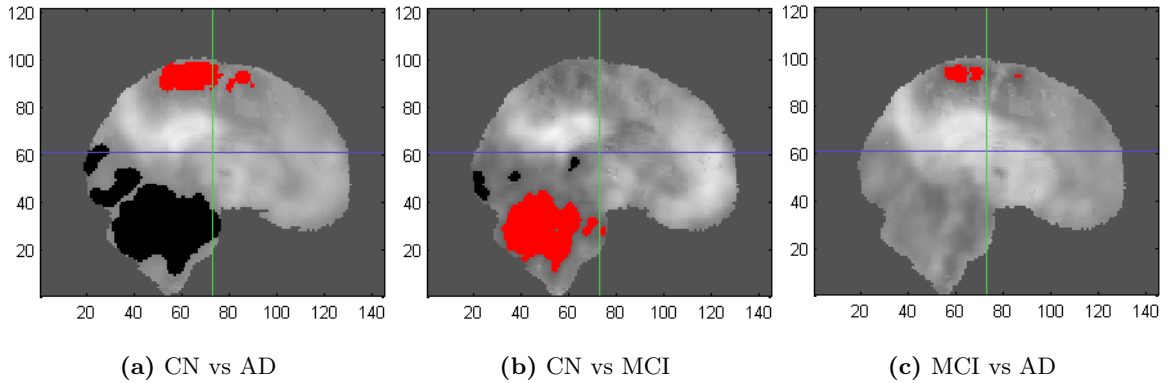
Figure 4.2: Transaxial section of the brain exhibiting part of the clusters obtained for (a) CN vs AD, (b) CN vs MCI and (c) MCI vs AD. The reference cluster is marked in red, and the remaining clusters extracted but not selected are marked in black.

Illustration of the mean of the FDG-PET baseline scans of the whole sample of CN and AD subjects after intensity normalization to CGM and reference cluster is presented in Fig. 4.3. Areas of apparent hypermetabolism (marked with a red circle) can be observed in AD subjects compared to CN subjects when CGM intensity normalization is performed. When the intensity of the FDG-PET images are normalized to the reference cluster, these artifacts are corrected. Consequently, the overall reduced metabolism in AD brain compared to CN brain becomes apparent, and the regions that before exhibited artificial hypermetabolism appear as conserved regions (marked with a green circle).

The classification results using the baseline data normalized to the CGM, regional mean (cerebellum and brainstem) and reference cluster are shown in Fig. 4.4. As one can observe, the reference cluster approach achieves higher accuracies compared to the other two methods, in all three classification tasks. However, the advantages of this method are not so marked for the MCI vs AD classification task, where the CGM approach reached accuracy values very close to the ones obtained with the
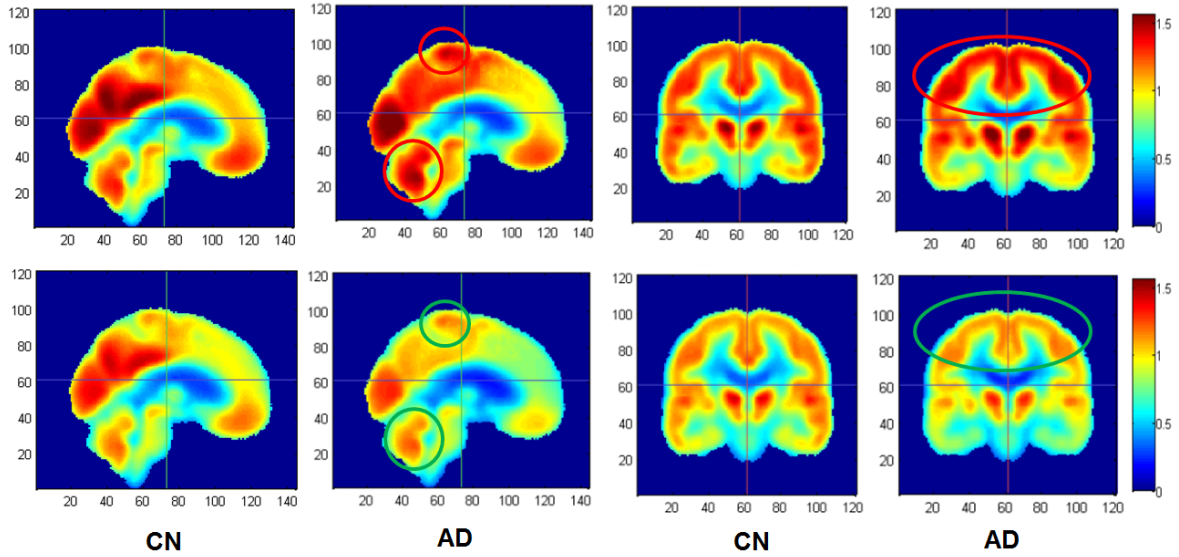
Figure 4.3: Illustration of the FGD-PET baseline scans (transaxial and coronal sections) of CN and AD subjects normalized to the GGM (top row) and reference cluster (bottom row). Red circles indicate hypermetabolic areas in AD subjects relative to the CN subjects. Green circles indicates the same regions after the reference cluster intensity normalization.

reference cluster. This could be due to the fact that for the reference cluster method, one of the two groups has to be defined as "normal". For the MCI vs AD classification, it is assumed that the MCI group is the control one, which is not accurate since patients diagnosed with MCI have already begun the neuronal degeneration process. On the other hand, the normalization performed relative to the brainstem appears to lead to a degradation of the classifier in CN vs MCI and MCI vs AD classification tasks. Since the reference cluster approach led to superior results, this method was adopted to normalize the baseline and follow-up imaging data.

4.3.2 Feature Selection Methods

First of all, the features selected by the t-test, linear correlation and mutual information were visually inspected. To this end, the whole sample of subjects were used and 3000 features were selected from the baseline scan and another 3000 features were selected from the VI differences along 24 months. No significant differences in the localization of the features selected by the three methods were detected using both sets of features. Hence, only the result of the t-test analysis is illustrated in Fig. 4.5, using the baseline data, and in Fig. 4.6, using the longitudinal data.

The selected regions in the baseline scans are mostly located in bilateral temporal lobe, frontal pole, posterior and anterior divisions of the parahippocampal gyrus and bilateral hippocampus between all groups. Additionally, the bilateral amygdala was also selected for CN vs AD and MCI vs AD. All these regions have been reported by several studies (for example, [51], [73], [74]) to be affected in both MCI and AD conditions. Furthermore, one can observe that higher t-values are obtained for CN vs AD which is related to higher metabolic differences between CN and AD patients. Lower t-values are obtained for CN vs MCI which reveals lower brain metabolic differences complicating the discrimination between CN and MCI subjects.

In the longitudinal data, the t-value is related to differences in metabolic change along time between groups. For CN vs AD and MCI vs AD, the features selected are very similar to the ones selected at

(a) CN vs AD
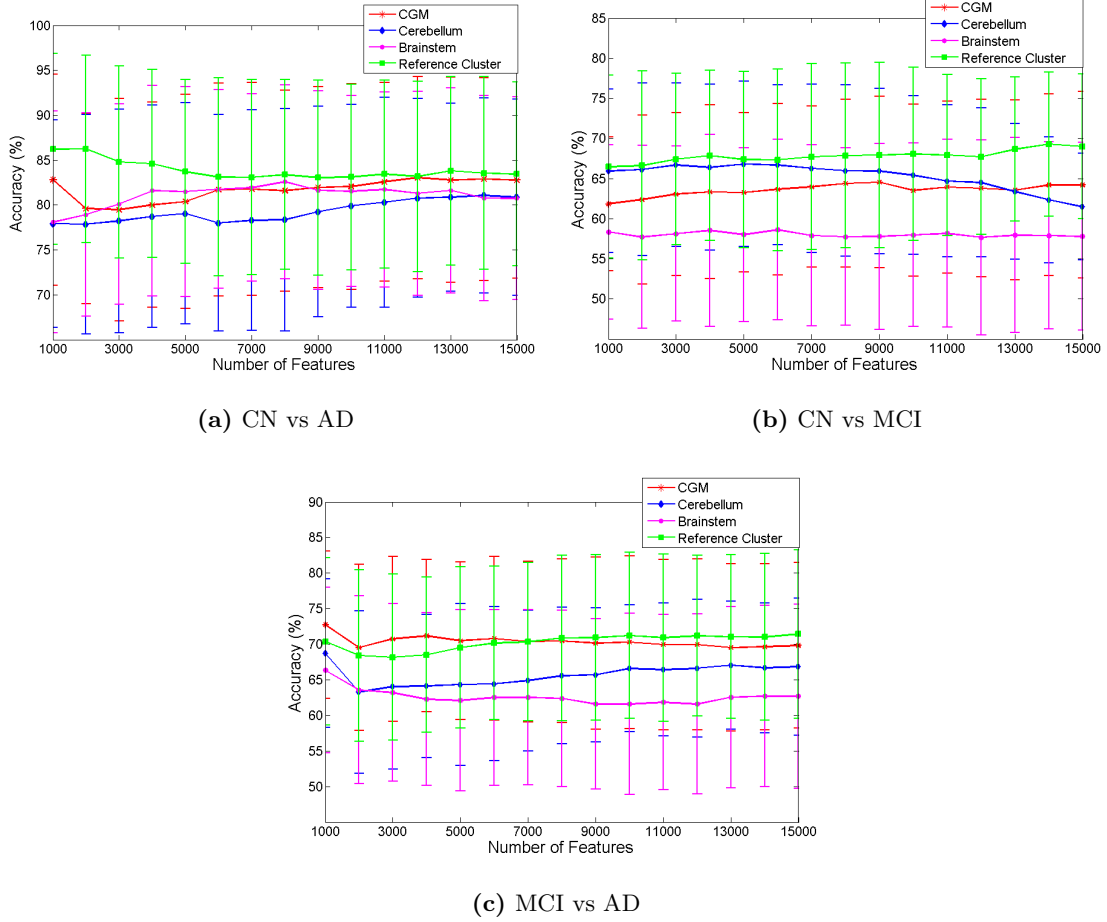

(b) CN vs MCI

(c) MCI vs AD

Figure 4.4: Classification results using baseline data normalized relative to the CGM, cerebellum, brainstem and reference cluster for (a) CN vs AD, (b) CN vs MCI, (c) MCI vs AD. The VI features of the baseline scans were selected with the t-test.

the baseline. However, for CN vs MCI, the selected features appear to have a more spread distribution across the brain. Note that during the classification process, the features are selected within the training set and not with the whole sample of subjects, therefore the features used for classification may vary, depending on the folds created in the cross-validation procedure.

The classification results using the t-test, the linear correlation and the mutual information as feature selection methods for the voxel-based analysis are presented in Fig. 4.7 using the baseline data, and in Fig. 4.8 using the VI change over 24 months.

For both sets of features, similar conclusions can be drawn. As one can observe in Fig. 4.7 and Fig. 4.8, although the mutual information method appears to lead to sightly higher accuracy results when using a larger number of features, overall, the classifier performance do not seem to benefit much from the increasing number of selected features. Note also that the minimum number of features used was 100 and the maximum was 300 times superior, and very similar accuracy results were obtained. This demonstrates the ability of SVM to circumvent the *curse of dimensionality*. Nevertheless, there are clear motivations to reduce the dimensionality of the input space, e.g., reduction the computational cost of the training and test algorithms, elimination of redundant information and selection of the most relevant features in order to gather more meaningful information from a physiological point of view.
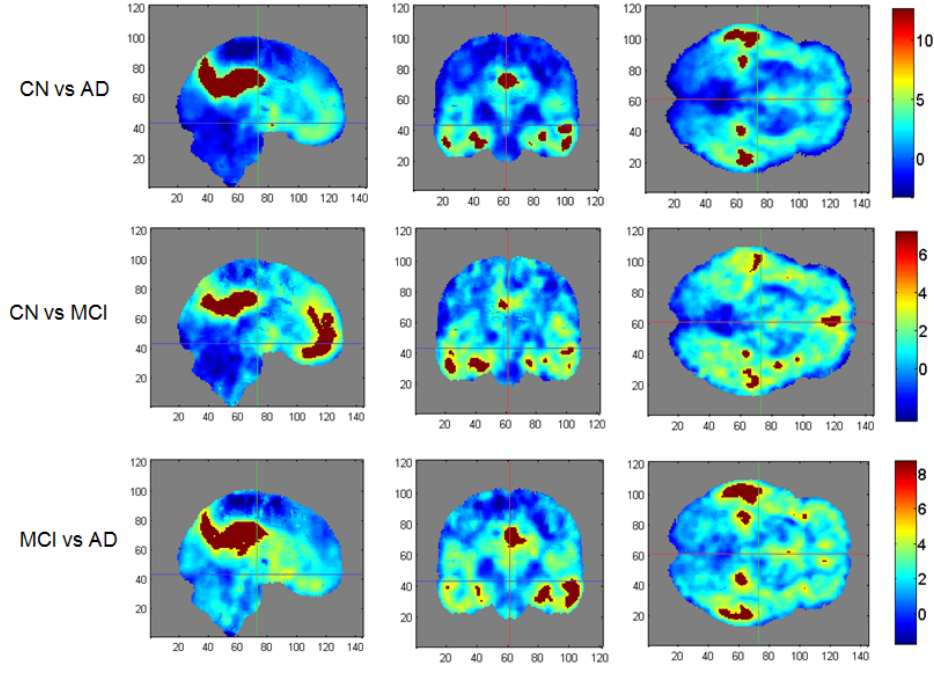
Figure 4.5: Differences in baseline glucose metabolism between CN, AD and MCI subjects detected with t-test. Sagital sections (left column), coronal sections (middle column) and transaxial sections (right column) of the FDG-PET scans are presented. Warmer colors correspond to higher t-values indicating higher metabolism differences between groups. The highlighted regions correspond to the 30000 selected voxels (voxels with higher t-values).
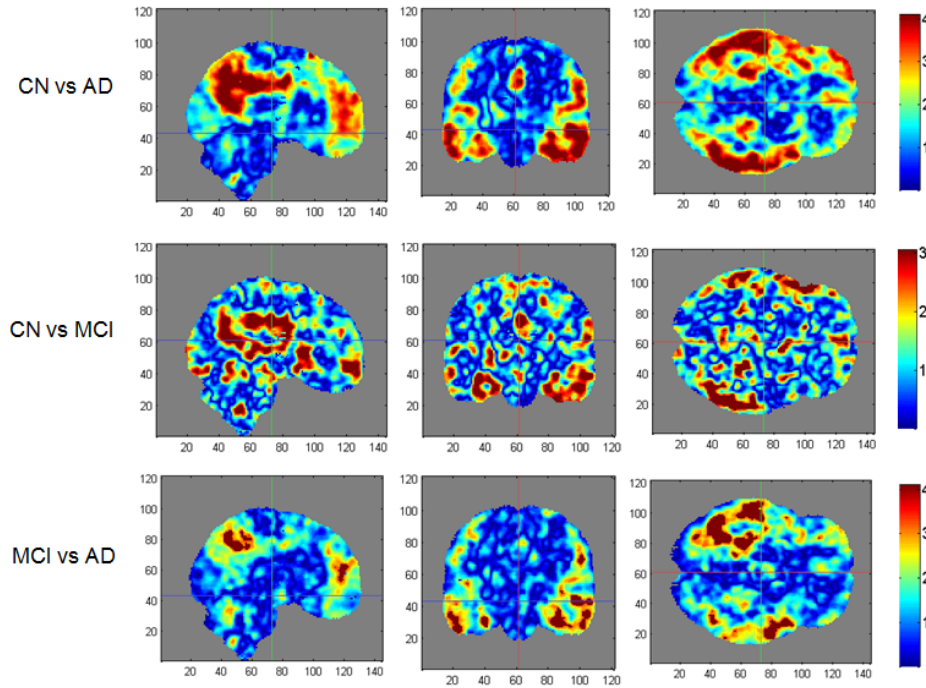

Figure 4.6: Differences in glucose metabolism over a 24 months period between CN, AD and MCI subjects detected with t-test. Sagital sections (left column), coronal sections (middle column) and transaxial sections (right column) of the FDG-PET scans are presented. Warmer colors correspond to higher t-values that indicate higher differences in metabolic change along time between groups. The highlighted regions correspond to the 30000 selected voxels (voxels with higher t-values).
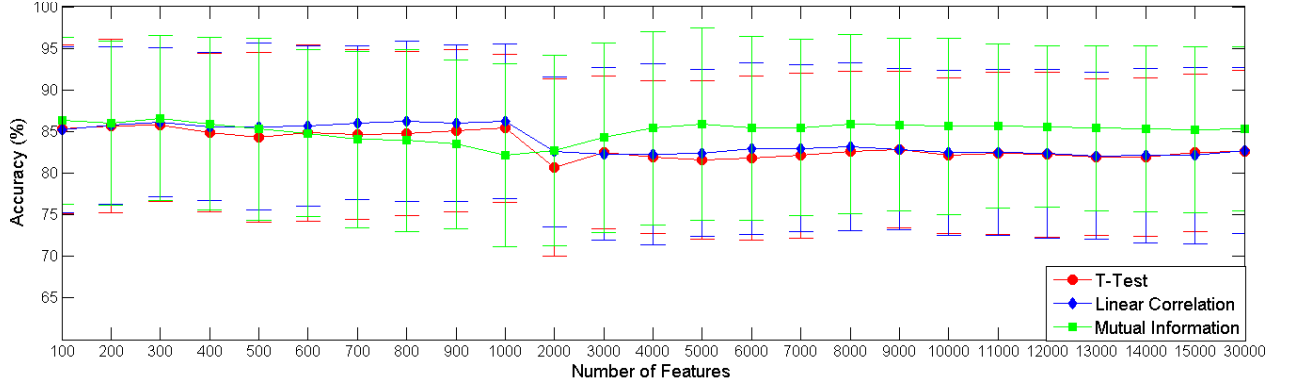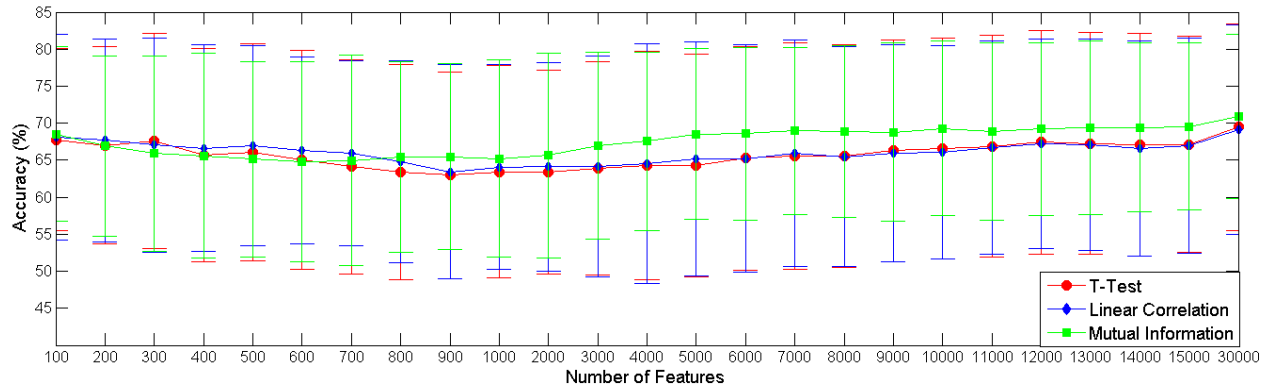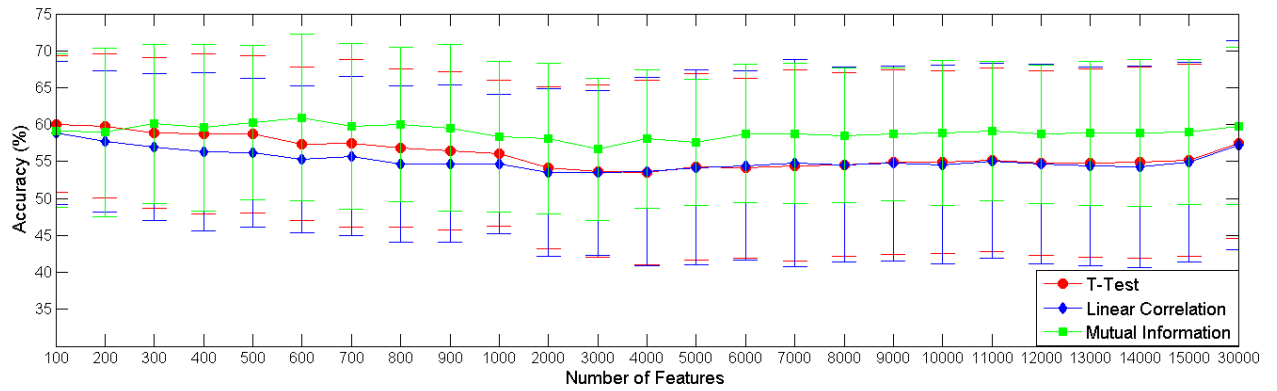
(a) CN vs AD
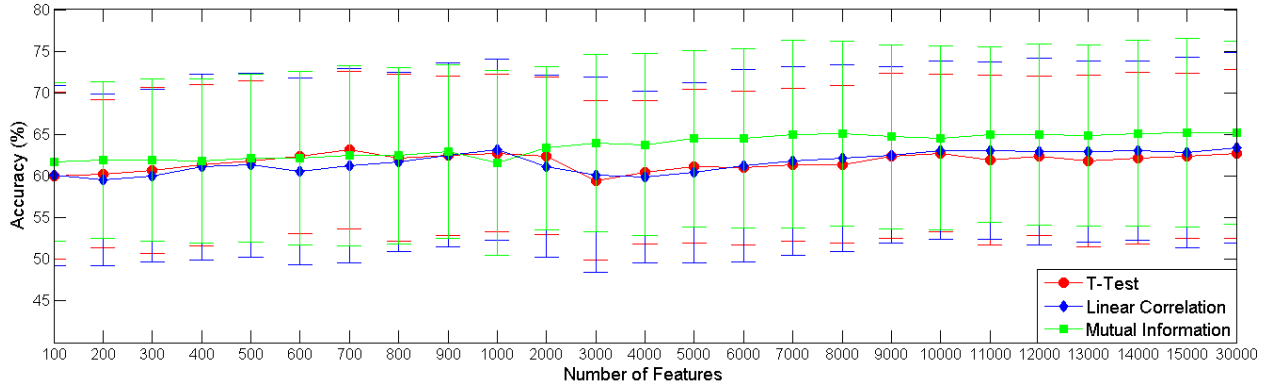
(b) CN vs MCI
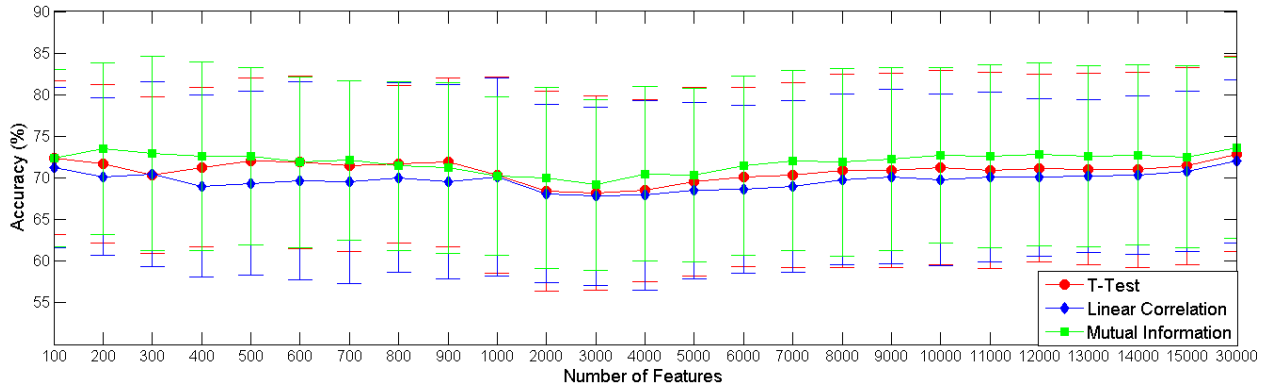
(c) MCI vs AD

Figure 4.7: Comparison of feature selection methods for varying the number of cross-sectional features in the voxel-based analysis. Mean accuracies and standard deviations are presented.

(a) CN vs AD

(b) CN vs MCI

(c) MCI vs AD

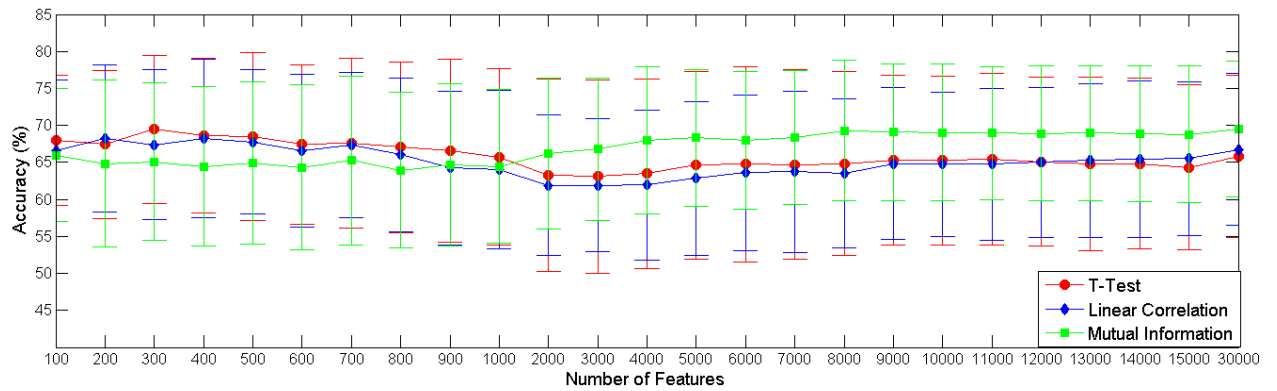**Figure 4.8:** Comparison of feature selection methods for varying the number of longitudinal features in the voxel-based analysis. Mean accuracies and standard deviations are presented.

The same classification results, but now for the multi-region analysis, are presented in Fig. 4.9 using the baseline scans, and in Fig. 4.10 using the regional change over 24 months. Similarly to the voxel-based analysis, none of the feature selection methods led to clearly superior accuracy results, indicating that the number of regions used did not have a great influence in the classification task.

Comparing the mean accuracies obtained with cross-sectional (Fig. 4.7 and Fig. 4.9) and longitudinal data (Fig. 4.8 and Fig. 4.10), one can conclude that the feature selection algorithms have better performance when using the cross-sectional data. In particular, for CN vs MCI using the VI change over 24 months as features, led to very poor classification performance. For this reason, in the following classification experiments, the features for longitudinal data were selected exclusively in the follow-up data, as explained in Section 4.2. For all classifications experiments, the CN vs AD classification achieved consistently the best results, whereas the CN vs MCI classification achieved lower mean accuracies in all classification tasks. These results were expected since the CMRgl differences between CN and AD subjects are larger than the differences between CN vs MCI.

Although mutual information appeared to be a more suitable method when a larger number of features is used, it has the drawback of computational cost. Furthermore, an evidence of accuracy improvement was not always observed in all the classification experiments when using the mutual information. Therefore, only the t-test was used as feature selection method in the following classification experiments.

### 4.3.3 Cross-Sectional and Longitudinal Classification Data Results

The classification accuracy results using the follow-up scans (6-month, 12-month and 24-month) and the corresponding differences relative to the baseline, for both voxel-based (using the best 1000 features selected with t-test) and multi-region analysis (using all 69 regions) are displayed in Fig. 4.11.

Similarly to the above observed results (Subsection 4.3.2), for CN vs AD classification (Fig. 4.11 (a)), the longitudinal data has consistently lower classifications results compared to cross-sectional data, for both analysis. Unlike what was expected, a similar performance for CN vs AD classification using a voxel-based analysis is observed between all the time-point scans and between the longitudinal data. Since AD is a progressive disorder, it was excepted that the metabolic differences along time between CN subjects would be more marked. However, these results may be explained by the fact that the feature selection was performed independently for each time-point. Hence, only the most discriminative features are selected in each follow-up imaging data. However, in the multi-region, the mean accuracy using the longitudinal and cross-sectional data tends to increase for larger temporal differences for CN vs AD classification. These results reflect the overall metabolic decline along time in AD patients compared to CN subjects. Since in the multi-region analysis, all 69 regions were used, it is possible to observe the expected increase in classification performance for larger temporal differences, unlike the voxel-based analysis for the reasons above mentioned.

For CN vs MCI classification (Fig. 4.11 (b)), the mean accuracy obtained using the cross-sectional and longitudinal data does not exhibit such obvious differences as in the former plot. In the multi-region approach, the best cross-sectional results were achieved for the 24-month, as expected. However, all the longitudinal sets had similar classification performances, reflecting the less marked differences of metabolic decline along time of CN and MCI subjects. For the voxel-based analysis, similar results
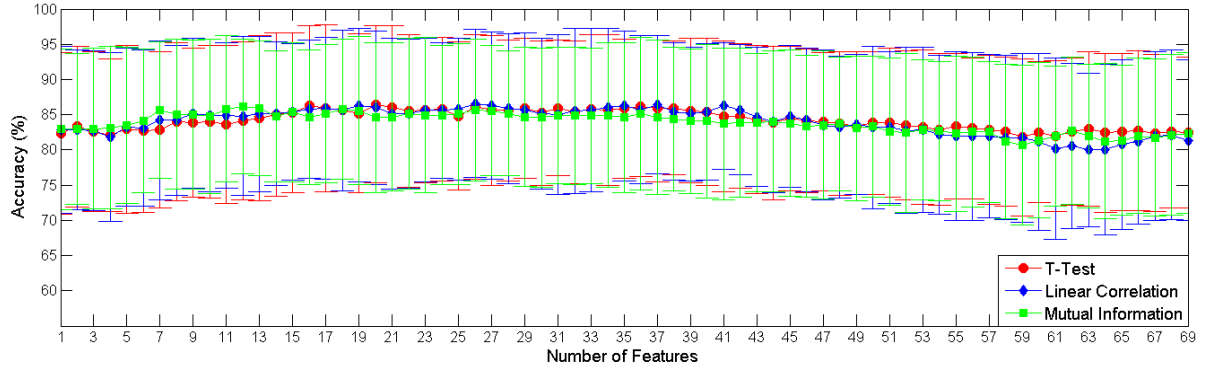
(a) CN vs AD
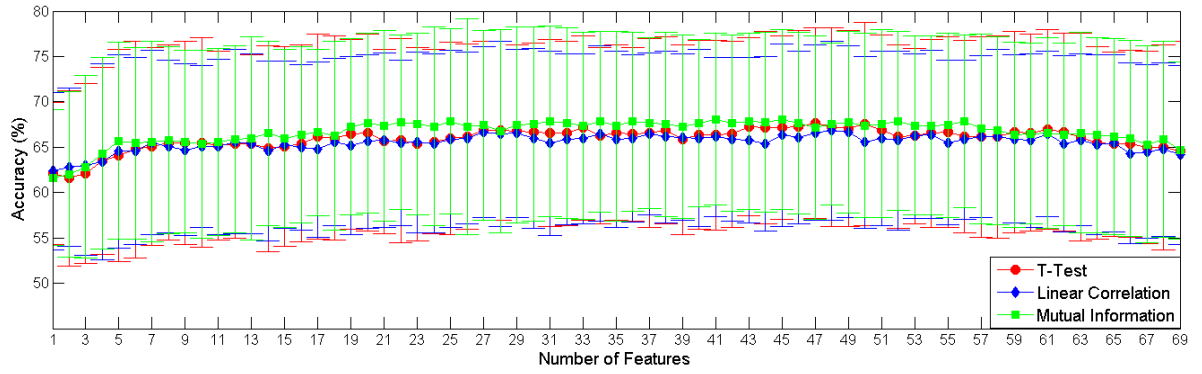
(b) CN vs MCI
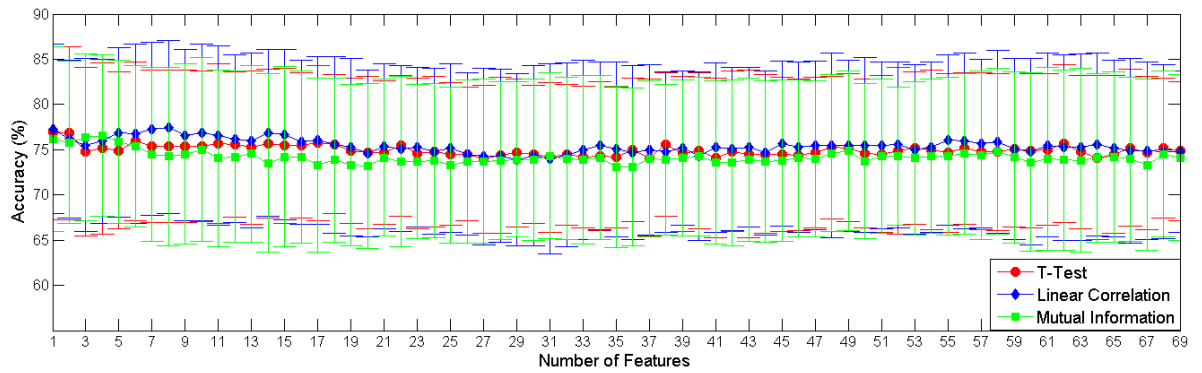
(c) MCI vs AD

Figure 4.9: Comparison of feature selection methods in the baseline data for varying the number of regional features. Mean accuracies and standard deviations are presented.
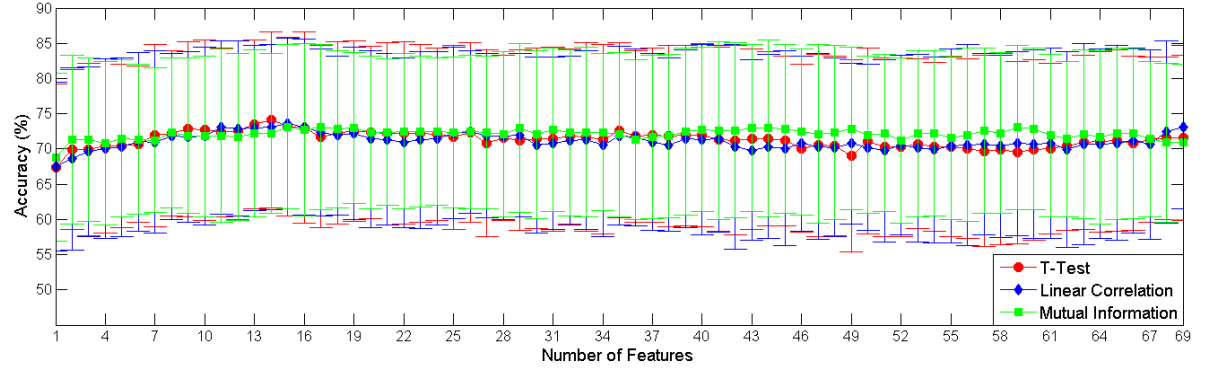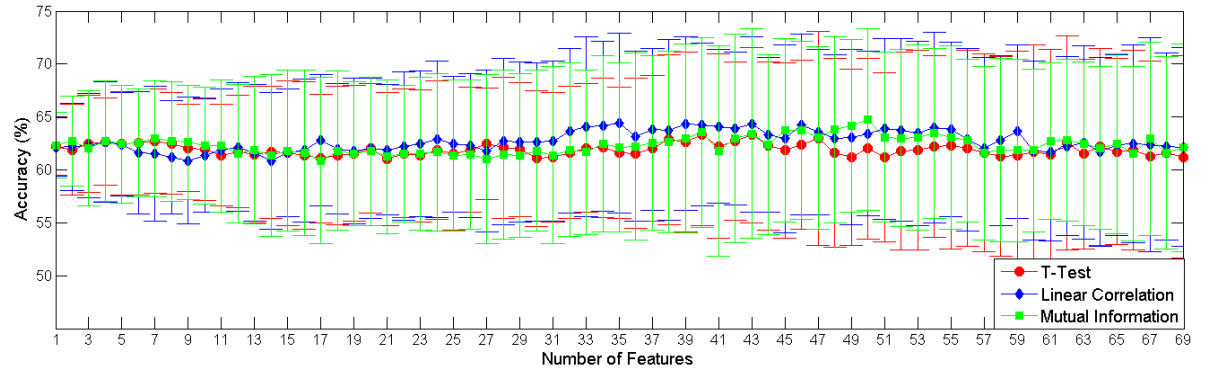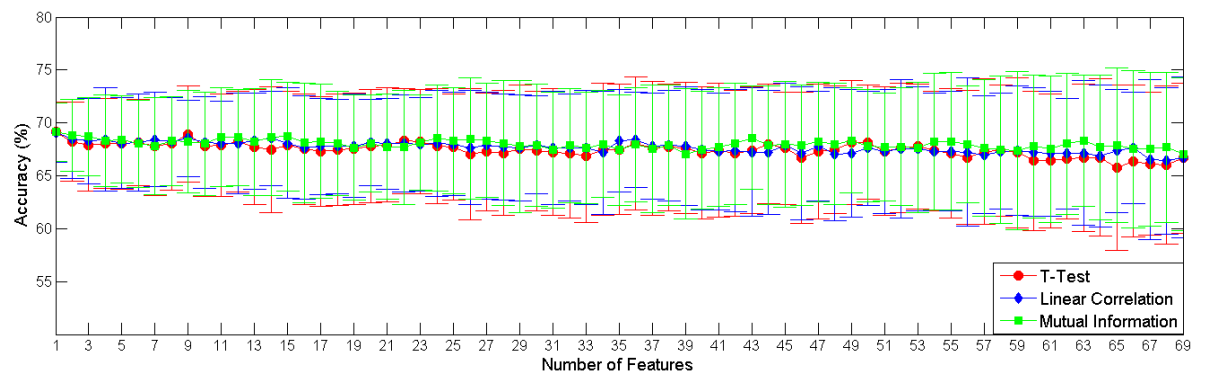
(a) CN vs AD

(b) CN vs MCI

(c) MCI vs AD

Figure 4.10: Comparison of feature selection methods in the longitudinal data for varying the number of regional features. Mean accuracies and standard deviations are presented.
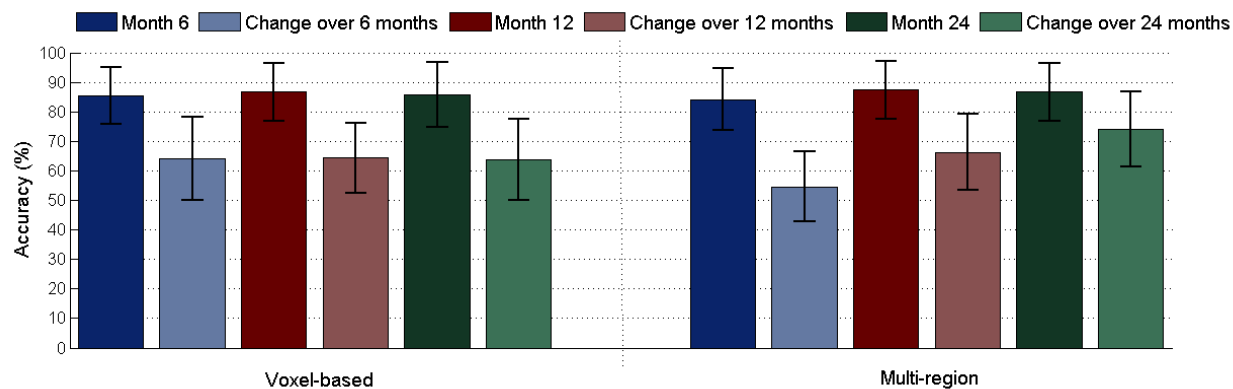
(a) CN vs AD
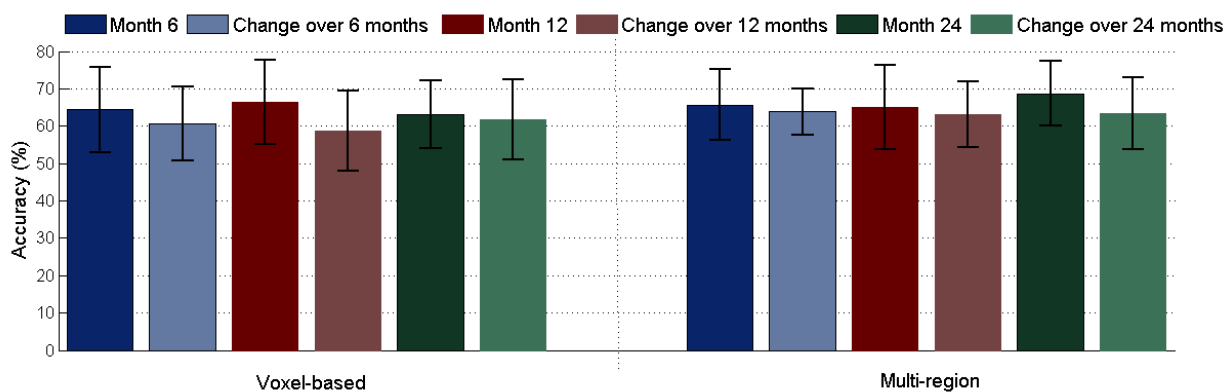
(b) CN vs MCI
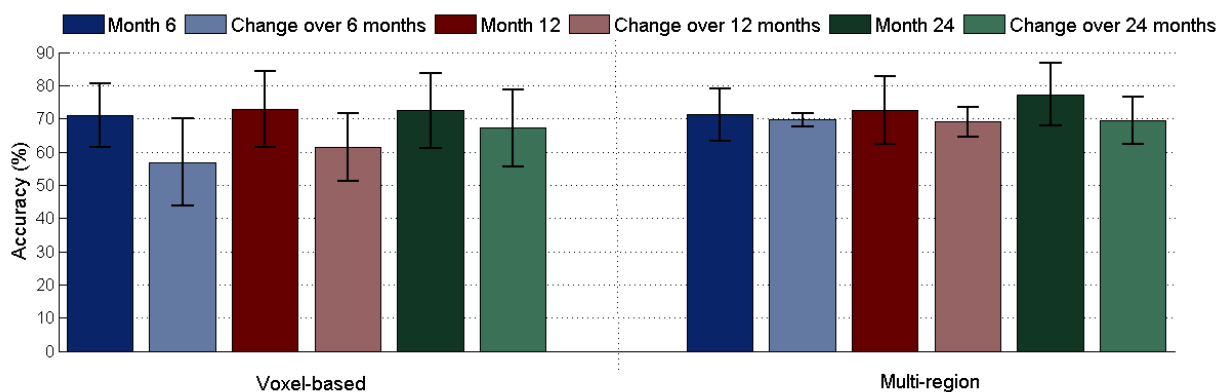
(c) MCI vs AD

Figure 4.11: Classification results using the follow-up data and the follow-up differences relative to the baseline, using a voxel-based (best 1000 features selected with t-test) and multi-region analysis (all 69 regions). Mean accuracies and standard deviations are presented.

were obtained between different cross-sectional data sets and between longitudinal data sets, which can be explained by the same reasons above mentioned for the CN vs AD classification.

For MCI vs AD (Fig. 4.11 (c)), both cross-sectional and longitudinal data achieved similar accuracies, except for the 24-month that outperformed the other sets of features when using regional features, similarly to CN vs MCI classification. For the voxel-based approach, there are evident differences between cross-sectional and longitudinal data. All the cross-sectional sets led to similar mean accuracies but classification performance of the cross-sectional data seemed do increase along time. For the 6-month change, the classification result is very poor, indicating that the metabolic decline over 6 months is not sufficient to discriminate between MCI and AD subjects even when using the best 1000 features. For the 12-month and 24-month change the classification results are in line with the results obtained for the other two classification tasks.

Table 4.3 presents the classification results obtained when combining cross-sectional and longitudinal data, for both voxel-based and multi-region analysis. Additionally, the results using only the cross-sectional data for all the single time point scans is also presented, in order to facilitate comparisons. Since the longitudinal data alone did not perform better than the cross-sectional data (Fig. 4.11), this information was not included in the table.

For the CN vs AD classification, the classification accuracy using the baseline data alone is significantly improved when combining the baseline data with the change over 12 months, and the baseline with change over 24 months, for both voxel-based and multi-region analysis. The best result obtained was a mean accuracy of 92.6 %, achieved for the combination of the 12-month data and the changes over 12 months using a voxel-based approach and 87.8 % using a multi-region approach. For the CN vs MCI classification, using a voxel-wise analysis, the combination of baseline and change over 12 months led do higher accuracies compared to using the baseline data alone. Moreover, the combination of month 6 with change over 6 months, and month 12 combined with the change over 12 months led to an improvement of the classification performances compared to using the cross-sectional data alone. The best results were achieved for the voxel-based approach, when concatenated the month 12 data with changes over 12 months (70.2 % against 65.6 % obtained with the multi-region analysis). For the MCI vs AD classification, using a voxel-based analysis, the combination of the baseline and change over 12 months led to an improvement of the mean accuracy. Moreover, the combination of all the follow-up data with the respective differences relative to baseline, led to an improvement of the accuracy. The most remarkable result was obtained for the combination of the month 12 with change over 12 months that achieved an accuracy of 69.5 %, against 65.7 % when using the cross-sectional data alone.

These results demonstrate that, although the longitudinal information alone did not have much discriminative power, when it is combined with the cross-sectional data, the classification performance is improved. However, the incorporation of longitudinal and cross-sectional data, in some particular cases, did not enhance the results. This can be due to the fact that cross-sectional and longitudinal features are chosen independently, and it may led to the incorporation of redundant information into the classifier.

In order to investigate the classifier performance for different number of VI features, the accuracy obtained when combining the baseline with change over 12 months was assessed using $F \in [1000:1000:15000]$. The results are displayed in Fig. 4.12 for the CN vs AD, CN vs MCI and MCI vs AD classification. In addition to the combination of baseline and longitudinal data, the accuracy

obtained using the baseline data alone is also presented in Fig. 4.12 in order to allow a direct comparison. Note that the number of features in each set, for the combination of cross-sectional and longitudinal data, is twice of the indicated in the axes, since both sets of features are concatenated. For all three classification experiments, the combination of longitudinal and cross-sectional information led to higher accuracies when compared to using only baseline data. Furthermore, these differences are more marked for CN vs AD. This result demonstrates that incorporating the information contained in longitudinal data, which reflects the larger metabolic decline in AD patients compared to CN subjects, led to a more successful discrimination between these two groups. The results of the t-test performed to the differences in the classification results obtained using these two sets of features, for all number of features, show that these are statistically different (p-value<0.05) for all three classification tasks.

	CN vs AD		CN vs MCI		MCI vs AD	
	Voxel-wise	Multi-region	Voxel-wise	Multi-region	Voxel-wise	Multi-region
Baseline	84.2 ± 11.1	81.1 ± 11.1	68.1 ± 10.6	68.5 ± 9.6	71.8 ± 10.3	75.7 ± 8.9
Baseline + change over 6 months	84.7 ± 10.2	81.2 ± 11.3	63.1 ± 11.2	68.4 ± 10.5	71.3 ± 10.4	72.6 ± 10.1
Baseline + change over 12 months	91.2 ± 8.0	83.3 ± 9.7	69.3 ± 10.9	$\mathbf{68.9 \pm 9.5}$	$\mathbf{73.1 \pm 10.5}$	72.3 ± 8.9
Baseline + change over 24 months	86.4 ± 9.5	87.7 ± 8.9	67.1 ± 9.8	66.9 ± 10.2	68.7 ± 11.0	73.5 ± 10.3
Month 6	86.6 ± 9.8	84.1 ± 10.5	62.0 ± 10.7	65.7 ± 9.5	65.8 ± 10.4	71.2 ± 7.9
Month 6 + change over 6 months	86.3 ± 9.6	83.5 ± 10.6	63.2 ± 12.4	65.2 ± 9.1	67.7 ± 10.2	71.7 ± 8.0
Month 12	$\mathbf{92.8 \pm 6.3}$	87.4 ± 9.8	69.7 ± 10.6	65.1 ± 11.3	65.7 ± 11.5	72.4 ± 10.2
Month 12 + change over 12 months	92.6 ± 6.7	$\mathbf{87.8 \pm 9.1}$	$\mathbf{70.2 \pm 9.0}$	65.6 ± 9.6	69.5 ± 11.0	72.7 ± 8.3
Month 24	86.9 ± 10.2	86.6 ± 9.9	66.0 ± 9.2	68.7 ± 8.7	69.1 ± 11.9	$\mathbf{77.3 \pm 9.0}$
Month 24 + change over 24 months	87.1 ± 9.8	87.3 ± 10.1	65.9 ± 8.7	66.7 ± 8.7	70.3 ± 9.7	74.2 ± 9.3

Table 4.3: Classification accuracy (mean ± standard deviation), performing a voxel-based and a multi-region analysis, for CN vs AD, CN vs MCI and MCI vs AD. The classification accuracy using the cross-sectional data alone (baseline, 6-month, 12-month, and 24-month) is presented, along with the classification accuracy obtained when combining the cross-sectional and the longitudinal data (change over 6, 12 and 24 months). The best result of each column is highlighted.
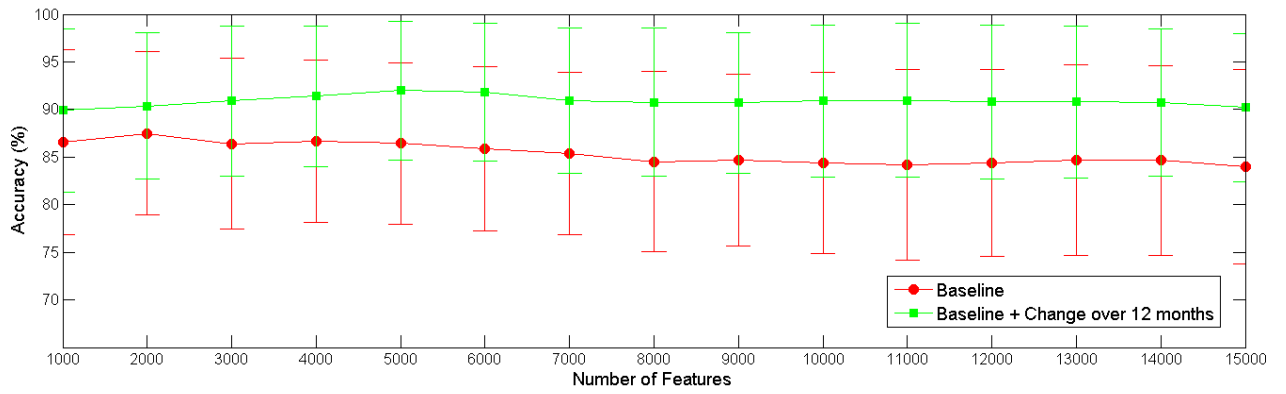
(a) CN vs AD
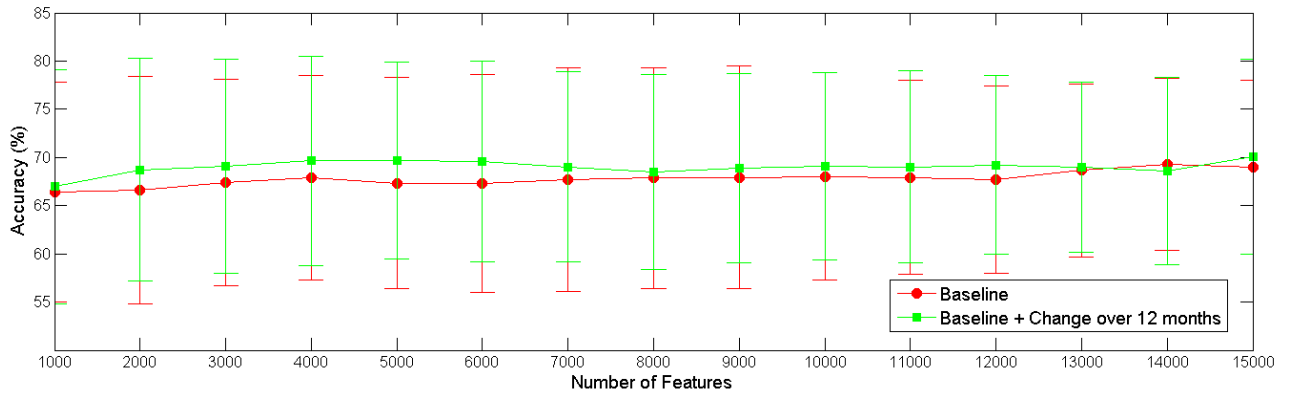
(b) CN vs MCI
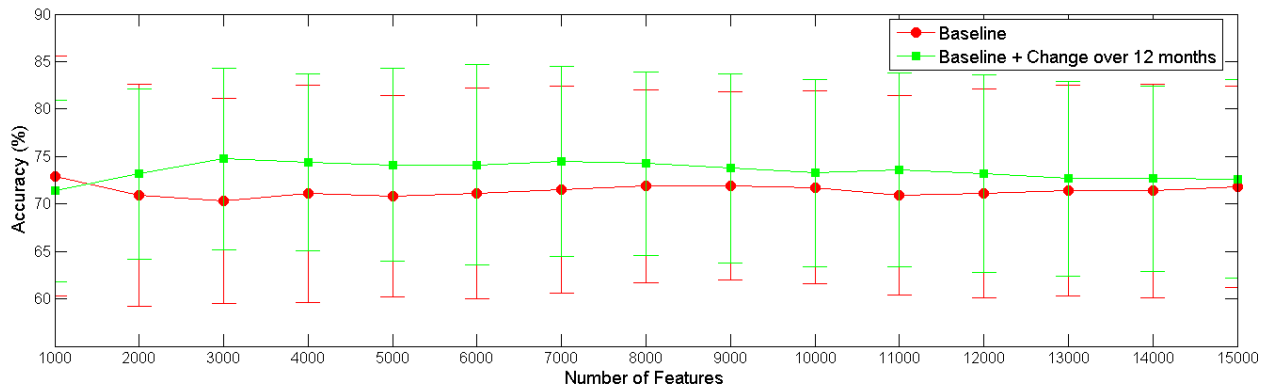
(c) MCI vs AD

Figure 4.12: Classification accuracy as a function of the number of features. The averaged accuracy and the standard deviation are presented.

5

Conclusions and Future Work

The present thesis aimed to investigate the value of longitudinal information for the automatic classification of AD using 3D FDG-PET images of the brain. The majority of the proposed CAD schemes for AD are based on the analysis of cross-sectional data. However, since AD is a progressive neurological disorder, the identification of metabolic decline patterns along time might be useful for early diagnosis and prediction of the disease course.

In the present work, a simple and intuitive approach was used to extract longitudinal information from the FDG-PET images. Although the information contained in the changes in metabolic pattern along time did not appear to have much discriminative power when used by itself, when is was used in combination with cross-sectional data it was able to enhance the classification performance. In particular, the combination of the baseline and VI change over 12 months led to significantly better accuracy results when compared to using only baseline data, for all the classification tasks and using different number of features. However, the incorporation of longitudinal and cross-sectional data did not lead to an improvement of the results, in some of the classification experiments. Since the cross-sectional and longitudinal features are chosen independently, the incorporation of redundant information into the classifier may occur, which can explain the deterioration of its performance. Nevertheless, even using simple feature extraction and selection methods, the value of longitudinal information for automatic diagnosis of AD was demonstrated. Furthermore, the results obtained are in line with the current state-of-the-art, achieving 91.2 % for CN vs AD, 69.3 % for CN vs MCI and 73.3 % for MCI vs AD, using the baseline data combined with change over 12 months, with a voxel-based approach. The results obtained in the present thesis suggest that the incorporation of longitudinal data in CAD systems for AD might provide valuable information for early diagnosis and prediction of the disease course.

The study of different approaches for intensity normalization of FDG-PET images conducted in the present work, was driven by the growing body of literature revealing a significant CGM reduction in AD patients compared to healthy controls. Therefore, global normalization, the most commonly used method, leads to artificial regional activity overestimating the true levels of CMRgl in AD patients compared to CN subjects. In the present study, it was proven that this phenomenon influences the classification task, since the features used are based on the CMRgl represented by the VI of the FDG-PET images. Among the intensity normalization approaches tested, the reference cluster method led to the best classification performance for all three classification tasks. On the other hand, the intensity normalization to cerebellum and brainstem, which have been widely reported in literature as conserved regions in AD condition, led to relatively poor classification performance, indicating that these regions may vary across different groups of subjects. Hence, these results suggest that a data-driven method is the most suitable approach to normalize the FDG-PET images, when comparing healthy controls and patients with neurological disorders. Moreover, since it is a data-driven method, the definition of a normalization region *a priori* is avoided, as well as the segmentation of the brain into ROIs.

Regarding the extraction of features from the FDG-PET images, a voxel-based and a multi-region analysis were conducted in present work. The voxel-based analysis had consistently better classification performances than the multi-region analysis for CN vs AD classification. For CN vs MCI classification task, both analysis led to similar results although, for MCI vs AD discrimination, the multi-region performed slightly better than the voxel-based approach. Even though it achieved lower accuracies for MCI vs AD classification, voxel-based analysis may still be a more advantageous option, since the seg-

mentation of the brain into ROIs is avoided. Furthermore, the pattern of degenerative brain disorders may not follow standard definitions for anatomical or functional regions, thus, brain segmentation into ROIs may lead to loss of discriminative information.

In order to overcome the *curse of dimensionality* emerged with the voxel-based approach, three feature selection methods were investigated, namely, the t-test, the linear correlation and the mutual information. For voxel-based analysis, both for cross-sectional and longitudinal data, the mutual information method performed slightly better in some of the classification tests when using a higher number of features, compared to two other methods. However, when a small number of features were used, none of the feature selection methods for the cross-sectional and longitudinal data led to clearly superior performances. In addition, the advantages of selecting the most discriminative regional features with these methods were also assessed. For both regional cross-sectional and longitudinal features, the results obtained did not evidence the benefits of feature selection procedure, and none of the methods outperformed the others. In the present thesis, the results obtained with mutual information did not outweigh its computational cost.

Overall, better classification results were obtained using the cross-sectional alone rather than the longitudinal data alone. However, these differences were less pronounced for CN vs MCI and MCI vs AD classification. In fact, when regional features were used, the results obtained with both set of features were very similar for these two classification tasks. These results suggest that the change of the metabolic pattern along time may provide valuable information for predicting the disease course in MCI subjects.

Despite these encouraging results, further investigation should be conducted in order to develop more sophisticated methods for longitudinal feature extraction and selection. Most existing feature selection methods are designed for a single time-point image and they can not be easily adapted to select features in multiple time-point images. Moreover, alternative strategies for the combination of the cross-sectional and longitudinal data more complex than the simple method of concatenation used in this work, should also be further explored. This procedure should eliminate redundant information and may improve the classification results when cross-sectional is used in combination with longitudinal data. In addition, a distinction between MCI converters and MCI non-converts might enhance the accuracy obtained between CN vs MCI and MCI vs AD classification tasks.

Bibliography

[1] K. Blennow, M. J. D. Leon, and H. Zetterberg, "Alzheimer' s disease." Lancet, 2006, vol. 368, pp. 387–403.

[2] J. Morris, "The Clinical Dementia Rating (CDR): current version and scoring rules." *Neurology*, vol. 43, no. November, pp. 2412–2414, 1993.

[3] M. Folstein, S. Folstein, and P. McHugh, ""Mini-mental state": a practical method for grading the cognitive state of patients for the clinician," *J. Psychiatr. Res.*, vol. 12, pp. 189–198, 1975.

[4] J. C. Morris, M. Storandt, J. P. Miller, D. W. McKeel, J. L. Price, E. H. Rubin, and L. Berg, "Mild cognitive impairment represents early-stage Alzheimer disease." *Arch. Neurol.*, vol. 58, no. 3, pp. 397–405, Mar. 2001.

[5] I. Mebane-Sims, "2013 Alzheimer's disease facts and figures." *Alzheimer's Dement.*, vol. 5, no. 3, pp. 234–70, Mar. 2013.

[6] C. Misra, Y. Fan, and C. Davatzikos, "Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI." *Neuroimage*, vol. 44, no. 4, pp. 1415–22, Feb. 2009.

[7] S. Eskildsen, P. Coupé, and D. García-Lorenzo, "Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning," *Neuroimage*, 2013.

[8] J. Gaugler, B. James, T. Johnson, K. Scholz, and J. Weuve, "2014 Alzheimer's disease facts and figures." Alzheimer's Association, Tech. Rep., 2014.

[9] J. Hardy and D. J. Selkoe, "The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics." *Science*, vol. 297, no. 5580, pp. 353–6, Jul. 2002.

[10] J. Raber, Y. Huang, and J. W. Ashford, "ApoE genotype accounts for the vast majority of AD risk and AD pathology." *Neurobiol. Aging*, vol. 25, no. 5, pp. 641–50, 2004.

[11] C. Ballard, S. Gauthier, A. Corbett, C. Brayne, D. Aarsland, and E. Jones, "Alzheimer's disease." *Lancet*, vol. 377, no. 9770, pp. 1019–31, Mar. 2011.

[12] I. Grundke-Iqbal, K. Iqbal, Y. C. Tung, M. Quinlan, H. M. Wisniewski, and L. I. Binder, "Abnormal phosphorylation of the microtubule-associated protein tau in Alzheimer cytoskeletal pathology." *Proc. Natl. Acad. Sci. U. S. A.*, vol. 83, no. 13, pp. 4913–7, Jul. 1986.

[13] K. Iqbal, A. D. C. Alonso, S. Chen, M. O. Chohan, E. El-Akkad, C.-X. Gong, S. Khatoon, B. Li, F. Liu, A. Rahman, H. Tanimukai, and I. Grundke-Iqbal, "Tau pathology in Alzheimer disease and other tauopathies." *Biochim. Biophys. Acta*, vol. 1739, no. 2-3, pp. 198–210, Jan. 2005.

[14] H. Braak, I. Alafuzoff, T. Arzberger, H. Kretzschmar, and K. Del Tredici, "Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry." *Acta Neuropathol.*, vol. 112, no. 4, pp. 389–404, Oct. 2006.

[15] S. a. Small and K. Duff, "Linking Abeta and tau in late-onset Alzheimer's disease: a dual pathway hypothesis." *Neuron*, vol. 60, no. 4, pp. 534–42, Nov. 2008.

[16] Ipsen, FondationY. Christen, R. M. Nitsch, C. Haass, R. M. Nitsch, K. Beyreuther and M. Jucker, *Alzheimer: 100 Years and Beyond.* Springer, 2006.

[17] M. Prince, M. Prina, and M. Guerchet, "World Alzheimer Report 2013," Alzheimer's Disease International, Tech. Rep., 2013.

[18] A. Wimo and M. Prince, "World Alzheimer Report 2010," *Alzheimer's Dis. Int.*, 2010.

[19] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub, and C. H. Phelps, "The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease." *Alzheimers. Dement.*, vol. 7, no. 3, pp. 263–9, May 2011.

[20] G. McKhann, D. Drachman, and M. Folstein, "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," Neurology, Tech. Rep., 1984.

[21] S. Basu, T. C. Kwee, S. Surti, E. a. Akin, D. Yoo, and A. Alavi, "Fundamentals of PET and PET/CT imaging." *Ann. N. Y. Acad. Sci.*, vol. 1228, pp. 1–18, Jun. 2011.

[22] F. Nobili and S. Morbelli, "18F-FDG-PET as a Biomarker for Early Alzheimer's Disease," *Open Nucl. Med. J.*, vol. 2, pp. 46–52, 2010.

[23] L. Sokoloff, "Relation between physiological function and energy metabolism in the central nervous system," *J. Neurochem.*, vol. 29, 1977.

[24] S. Minoshima, B. Giordani, S. Berent, K. a. Frey, N. L. Foster, and D. E. Kuhl, "Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease," *Ann. Neurol.*, vol. 42, no. 1, pp. 85–94, Jul. 1997.

[25] K. Herholz, E. Salmon, D. Perani, J. C. Baron, V. Holthoff, L. Frölich, P. Schönknecht, K. Ito, R. Mielke, E. Kalbe, G. Zündorf, X. Delbeuck, O. Pelati, D. Anchisi, F. Fazio, N. Kerrouche, B. Desgranges, F. Eustache, B. Beuthien-Baumann, C. Menzel, J. Schröder, T. Kato, Y. Arahata, M. Henze, and W. D. Heiss, "Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG-PET." *Neuroimage*, vol. 17, no. 1, pp. 302–16, Sep. 2002.

[26] L. Mosconi, W. H. Tsui, A. Pupi, S. De Santi, A. Drzezga, S. Minoshima, and M. J. de Leon, "(18)F-FDG PET database of longitudinally confirmed healthy elderly individuals improves detection of mild cognitive impairment and Alzheimer's disease." *J. Nucl. Med.*, vol. 48, no. 7, pp. 1129–34, Jul. 2007.

[27] K. Chen, J. B. S. Langbaum, A. S. Fleisher, N. Ayutyanont, C. Reschke, W. Lee, X. Liu, D. Bandy, G. E. Alexander, P. M. Thompson, N. L. Foster, D. J. Harvey, M. J. de Leon, R. a. Koeppe, W. J. Jagust, M. W. Weiner, and E. M. Reiman, "Twelve-month metabolic declines in probable Alzheimer's disease and amnestic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the Alzheimer's Disease Neuroimaging Initiative." *Neuroimage*, vol. 51, no. 2, pp. 654–64, Jun. 2010.

[28] I. Illán, J. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, F. Segovia, R. Chaves, M. Gómez-Rio, and C. Puntonet, "18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis," *Inf. Sci. (Ny).*, vol. 181, no. 4, pp. 903–916, Feb. 2011.

[29] M. Ewers, M. Brendel, A. Rizk-Jackson, A. Rominger, P. Bartenstein, N. Schuff, and M. W. Weiner, "Reduced FDG-PET brain metabolism and executive function predict clinical progression in elderly healthy subjects." *NeuroImage Clin.*, vol. 4, pp. 45–52, Jan. 2013.

[30] G. W. Small, S. Y. Bookheimer, P. M. Thompson, G. M. Cole, S.-C. Huang, V. Kepe, and J. R. Barrio, "Current and future uses of neuroimaging for cognitively impaired patients." *Lancet Neurol.*, vol. 7, no. 2, pp. 161–72, Feb. 2008.

[31] C. Hinrichs, V. Singh, G. Xu, and S. C. Johnson, "Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population." *Neuroimage*, vol. 55, no. 2, pp. 574–89, Mar. 2011.

[32] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment." *Neuroimage*, vol. 55, no. 3, pp. 856–67, Apr. 2011.

[33] K. R. Gray, R. Wolz, R. Heckemann, P. Aljabar, A. Hammers, and D. Rueckert, "Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease." *Neuroimage*, vol. 60, no. 1, pp. 221–9, Mar. 2012.

[34] I. Yakushev, A. Hammers, A. Fellgiebel, I. Schmidtmann, A. Scheurich, H.-G. Buchholz, J. Peters, P. Bartenstein, K. Lieb, and M. Schreckenberger, "SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnestic mild cognitive impairment from healthy aging." *Neuroimage*, vol. 44, no. 1, pp. 43–50, Jan. 2009.

[35] P. Borghammer, J. Aanerud, and A. Gjedde, "Data-driven intensity normalization of PET group comparison studies is superior to global mean normalization." *Neuroimage*, vol. 46, no. 4, pp. 981–8, Jul. 2009.

[36] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Comput. Med. imaging Graph.*, vol. 31, pp. 198–211, 2007.

[37] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. a. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, E. Liu, J. C. Morris, R. C. Petersen, A. J. Saykin, M. E. Schmidt, L. Shaw, L. Shen, J. a. Siuciak, H. Soares, A. W. Toga, and J. Q. Trojanowski, "The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception." *Alzheimers. Dement.*, vol. 9, no. 5, pp. e111–94, Sep. 2013.

[38] G. Chetelat and J.-C. Baron, "Early diagnosis of alzheimer's disease: contribution of structural neuroimaging," *Neuroimage*, vol. 18, no. 2, pp. 525–541, Feb. 2003.

[39] J. Stoeckel, G. Malandain, O. Migneco, P. M. Koulibaly, P. Robert, N. Ayache, and J. Darcourt, "Classification of SPECT images of normal subjects versus images of Alzheimer's disease patients," in *4th Int. Conf. Med. Image Comput. Comput. Interv.*, 2001, pp. 666–674.

[40] J. Stoeckel and G. Fung, "SVM Feature Selection for Classification of SPECT Images of Alzheimer's Disease using Spatial Information Computer Aided Diagnosis Computer Aided Diagnosis," in *Proc. Fifth IEEE Int. Conf. Data Min.* Washington, DC, USA: IEEE Computer Society, 2005.

[41] J. Stoeckel, N. Ayache, G. Malandain, P. M. Koulibaly, K. P. Ebmeier, and J. Darcourt, "Automatic classification of SPECT images of Alzheimer's disease patients and control subjects," *Med. Image . . .*, pp. 654–662, 2004.

[42] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick, "Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging." *Neurobiol. Aging*, vol. 29, no. 4, pp. 514–23, Apr. 2006.

[43] O. Colliot, G. Chételat, and M. Chupin, "Discrimination between Alzheimer Disease, Mild Cognitive Impairment, and Normal Aging by Using Automated Segmentation of the Hippocampus1," *Radiology*, vol. 248, no. 1, 2008.

[44] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy, and H. Benali, "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI." *Neuroradiology*, vol. 51, no. 2, pp. 73–83, Feb. 2009.

[45] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database." *Neuroimage*, vol. 56, no. 2, pp. 766–81, May 2010.

[46] F. Liu, C.-Y. Wee, H. Chen, and D. Shen, "Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification." *Neuroimage*, vol. 84C, pp. 466–475, Sep. 2013.

[47] J. Dukart, K. Mueller, A. Horstmann, B. Vogt, S. Frisch, H. Barthel, G. Becker, H. E. Möller, A. Villringer, O. Sabri, and M. L. Schroeter, "Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies." *Neuroimage*, vol. 49, no. 2, pp. 1490–5, Jan. 2010.

[48] A. Küntzelmann, T. Guenther, U. Haberkorn, M. Essig, F. Giesel, R. Henze, M. L. Schroeter, J. Schröder, and P. Schönknecht, "Impaired cerebral glucose metabolism in prodromal Alzheimer's disease differs by regional intensity normalization." *Neurosci. Lett.*, vol. 534, pp. 12–7, Feb. 2013.

[49] J. B. S. Langbaum, K. Chen, W. Lee, C. Reschke, D. Bandy, A. S. Fleisher, G. E. Alexander, N. L. Foster, M. W. Weiner, R. a. Koeppe, W. J. Jagust, and E. M. Reiman, "Categorical and correlational analyses of baseline fluorodeoxyglucose positron emission tomography images from the Alzheimer's Disease Neuroimaging Initiative (ADNI)." *Neuroimage*, vol. 45, no. 4, pp. 1107–16, May 2009.

[50] J. L. Andersson, "How to estimate global activity independent of changes in local activity." *Neuroimage*, vol. 6, no. 4, pp. 237–44, Nov. 1997.

[51] G. Sanabria-Diaz, E. Martínez-Montes, and L. Melie-Garcia, "Glucose metabolism during resting state reveals abnormal brain networks organization in the Alzheimer's disease and mild cognitive impairment." *PLoS One*, vol. 8, no. 7, p. e68860, Jan. 2013.

[52] J. Dukart, R. Perneczky, S. Förster, H. Barthel, J. Diehl-Schmid, B. Draganski, H. Obrig, E. Santarnecchi, A. Drzezga, A. Fellgiebel, R. Frackowiak, A. Kurz, K. Müller, O. Sabri, M. L. Schroeter, and I. Yakushev, "Reference cluster normalization improves detection of frontotemporal lobar degeneration by means of FDG-PET." *PLoS One*, vol. 8, no. 2, p. e55415, Jan. 2013.

[53] V. Berti, C. Polito, P. Borghammer, S. Ramat, L. Mosconi, E. Vanzi, M. T. D. Cristofaro, M. D. Leon, S. Sorbi, and A. Pupi, "Alternative normalization methods demonstrate widespread cortical hypometabolism in untreated de novo Parkinson's disease," *Q J Nucl Med Mol Imaging*, vol. 56, no. 3, pp. 299–308, 2012.

[54] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale, "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain." *Neuron*, vol. 33, no. 3, pp. 341–55, Jan. 2002.

[55] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest." *Neuroimage*, vol. 31, no. 3, pp. 968–80, Jul. 2006.

[56] O. T. Carmichael, H. a. Aizenstein, S. W. Davis, J. T. Becker, P. M. Thompson, C. C. Meltzer, and Y. Liu, "Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment." *Neuroimage*, vol. 27, no. 4, pp. 979–90, Oct. 2005.

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

[58] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[59] L. Li, J. Wang, and D. Chahal, "Detection of Mild Cognitive Impairment Using Image Differences and Clinical Features," *IEEE International Conference on Bioinformatics and Bioengineering Detection*, pp. 106–111, 2010.

[60] R. Battiti, "Using mutual information for selecting features in supervised neural net learning." *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 537–50, Jan. 1994.

[61] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. ACM Work. Comput. Learn. Theory*, A. Press, Ed., 1992.

[62] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 7, 1995.

[63] J. Ramírez, J. Górriz, D. Salas-Gonzalez, A. Romero, M. López, I. Álvarez, and M. Gómez-Río, "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features," *Inf. Sci. (Ny).*, vol. 237, pp. 59–72, Jul. 2013.

[64] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.

[65] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection." *BMC Bioinformatics*, vol. 7, p. 91, Jan. 2006.

[66] "Alzheimer's Disease Neuroimaging Initiative - PET Technical Procedures Manual," http://adni.loni.usc.edu/methods/pet-analysis/, online: acessed March 1, 2014.

[67] W. J. Jagust, D. Bandy, K. Chen, N. L. Foster, S. M. Landau, C. a. Mathis, J. C. Price, E. M. Reiman, D. Skovronsky, and R. a. Koeppe, "The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core." *Alzheimers. Dement.*, vol. 6, no. 3, pp. 221–9, May 2010.

[68] F. Ségonne, a. M. Dale, E. Busa, M. Glessner, D. Salat, H. K. Hahn, and B. Fischl, "A hybrid approach to the skull stripping problem in MRI." *Neuroimage*, vol. 22, no. 3, pp. 1060–75, Jul. 2004.

[69] J. Ashburner and K. J. Friston, "Unified segmentation," *Neuroimage*, vol. 26, no. 3, pp. 839–51, Jul. 2005.

[70] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," *Inf. Process. Med. imaging*, vol. 3, 1995.

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

[72] V. S. Fonov, A. C. Evans, R. C. McKinstry, C. R. Almli, and D. L. Collins, "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood," *Neuroimage*, vol. 47:S102, 2009.

[73] S. De Santi, M. J. de Leon, H. Rusinek, a. Convit, C. Y. Tarshish, a. Roche, W. H. Tsui, E. Kandil, M. Boppana, K. Daisley, G. J. Wang, D. Schlyer, and J. Fowler, "Hippocampal formation glucose metabolism and volume losses in MCI and AD." *Neurobiol. Aging*, vol. 22, no. 4, pp. 529–39, 2001.

[74] D. Mitchell and C. Turkington, "The Encyclopedia of Alzheimer's Disease," 2010.