Official Doctoral Programme in Information Technology and Communication

Department of Signal Theory, Telematics and Communications

University of Granada

PhD THESIS

Novel approaches based on decomposition methods for detecting MRI patterns in the progression of Alzheimer's Disease

Written by:

Laila Khedher

Supervised by:

Dr. Javier Ramírez Pérez de Inestrosa

Dr. Dr. Juan Manuel Górriz Sáez

Editor: Universidad de Granada. Tesis Doctorales
Autora: Laila Khedher
ISBN: 978-84-9163-355-6
URI: http://hdl.handle.net/10481/47618

Programa Oficial de Doctorado en Tecnologías de la Información y la Comunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones

Universidad de Granada

TESIS DOCTORAL

Nuevas técnicas de análisis de imágenes de resonancia magnética para determinación de patrones de atrofia cerebral en la progresión de la enfermedad de Alzheimer

Realizado por:

Laila Khedher

Dirigido por:

Dr. Javier Ramírez Pérez de Inestrosa

Dr. Dr. Juan Manuel Górriz Sáez

La doctoranda Laila Khedher y los directores de la tesis Javier Ramírez Pérez de Inestrosa y Juan Manuel Górriz Sáez Garantizamos, al firmar esta tesis de los doctoral, que el trabajo ha sido realizado por la doctoranda bajo la dirección directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, Julio de 2017

Directores de la Tesis

Dr. Javier Ramírez Pérez de Inestrosa Dr. Dr. Juan Manuel Górriz Sáez

Doctoranda

Laila Khedher

"The secret of change is to focus all of your energy, not on fighting the old, but on building the new"
SOCRATES

Acknowledgements

Foremost I would like to thank Dr. Javier Ramírez Pérez de Inestrosa and Dr. Dr. Juan Manuel Górriz Sáez for supervising this work, their valuable comments and suggestions and specially for their continuous support, motivation and friendship.

I would also like to express my gratitude to Dr. Carlos García Puntonet, Dr. Ignacío Alvárez Illán, Dr. F. Segovia, Dr. Francisco Jesús Martínez Murcia, Dr. Pablo Padilla and Dr. Diego Salas-Gonzalez for the fruitful discussions.
Thank you Dr. Abdelbasset Brahim for carefully reading this thesis. His reading of the thesis has certainly improved its quality.

Some special words of gratitude go to the members of the Signal Processing and Biomedical Applications (SIPBA), who are a role model for me to become a successful researcher, for being such a nice group of people.

A very special word of thanks goes for my family for their relentless moral assistance during this Ph.D years.

Finally, I have to thank my husband and love of my life for keeping things going and for always showing how proud he is of me.

The last word goes for YASMINA, my baby girl, who has been the light of my life for the last few months and who has given me the extra strength and motivation to get things done. This thesis is dedicated to her.

Laila Khedher

Granada, July 2017

Abstract

Image-based computer aided diagnosis (CAD) systems have significant potential for screening and early detection of brain diseases. In this sense, this PhD work is motivated by the development and the implementation of novel CAD systems based on several pattern recognition/classification techniques for the early detection of neurodegenerative diseases. In particular, the dissertation is focused on the analysis of the most relevant one, the Alzheimer's disease (AD), by the use of structural magnetic resonance imaging (sMRI) techniques. The proposed CAD systems are based on several processing steps including segmentation of brain tissues (gray matter and white matter tissues), feature selection techniques such as t-test model, feature extraction techniques such as; Partial least squares (PLS), Principal Component Analysis (PCA), Independent component analysis (ICA) and Non-Negative Matrix Factorization (NNMF) techniques, and an automatic classification technique, such as the support vector machine (SVM). Most of these thesis contributions are included within the feature extraction techniques and its application to the development of automatic CAD systems for early detection of AD.

The first proposed CAD system is based on the PLS approach that extracts the relevant features to characterize the AD pattern. This technique decomposes two sets of variables into the product of two matrices called *scores* and *loadings* according to a criterion of covariance maximization. In this work, these variables are those formed by the structural magnetic resonance images under study and the labels of these images. After the decomposition of these sets, the *scores* are used as feature vectors for the classification step.

The second proposed CAD system for AD detection is based on the PCA approach, as a feature extraction technique for sMRI brain images. This approach reduces the original high-dimensional space of the brain images to a lower dimensional subspace. PCA generates a set of orthonormal basis vectors, known as Principal components (PCs), that maximizes the scatter of all the projected samples, which is equivalent to diagonalize the covariance matrix through the eigenvalue of these components.

The third CAD system is based on the Independent component analysis (ICA) approach. At the beginning, a "template" image is computed as the average of healthy subject images or as the difference between normal and pathological images. Then, the ICA algorithm is applied to extract the maximally spatially independent components (ICs) revealing patterns of variation that occur in the dataset under study.

The last method proposed is based on the NNMF approach, as a useful decomposition technique of multivariate data that solves the problem of finding non-negative matrices. These feature extraction techniques successfully solved the small sample size problem by obtaining only the relevant information related to AD. This process is known as a dimensionality reduction and it improves the prediction accuracy of CAD systems, specifically, in the early stage of AD. In this way, considering the fact

that an early diagnosis of AD is crucial, classification experiments were performed not only to distinguish Normal Control (NC) and AD subjects but also to differentiate NC from a transitional phase between being cognitively normal and having an AD diagnosis. This later phase is called Mild Cognitive Impairment (MCI).
The proposed feature extraction methods have been combined with SVM classifiers and the accuracy rates of the resulting CAD systems have been estimated by means of a sMRI database from the Alzheimer disease neuroimaging initiative (ADNI). Furthermore, a kfold-cross validation technique was applied to these systems in order to tune the classifier parameters and to estimate its performance. The obtained results demonstrate the effectiveness and the robustness of the proposed CAD systems (with accuracy value around 90%) compared to previous approaches such as the one based on Voxel-As-Features (VAF) technique.

Resumen

Los sistemas de diagnóstico asistido por ordenador (CAD, del inglés "Computer Aided Diagnosis") basados en imágenes tienen un potencial significativo para el diagnóstico precoz y prognosis de enfermedades neurológicas. En este sentido, este trabajo de doctorado está motivado por el desarrollo y la implementación de nuevos sistemas CAD basados en varias técnicas de reconocimiento/clasificación de patrones para la detección temprana de enfermedades neurodegenerativas. En particular, la tesis se centra en el análisis de la más relevante, la enfermedad de Alzheimer (EA), mediante el uso de técnicas de resonancia magnética estructural (MRI, del inglés "magnetic resonance imaging").

Los sistemas CAD propuestos se basan en varios pasos de procesamiento incluyendo segmentación, selección de características, extracción de características y clasificación. Por ejemplo, en el paso de segmentación, el software SPM se utiliza para separar los tejidos cerebrales en materia gris, materia blanca y CSF. Sin embargo, nos preocupamos solamente en la materia gris y la materia blanca, ya que son las regiones más afectadas en la EA. Además, se ha utilizado una técnica de selección de características basada en un test "t" para seleccionar la intensidad del voxel significativa en las regiones de interés (ROI). Respecto a la etapa de extracción de características, el primer sistema CAD propuesto se basa en la técnica de los mínimos cuadrados parciales (PLS, del inglés "Partial least squares") para la extracción de las características relevantes que caracterizan la EA. Esta técnica transforma dos conjuntos de variables en el producto de dos matrices llamadas "*scores*" y "*loadings*" siguiendo un criterio de maximización de la covarianza. En este trabajo de doctorado, los dos conjuntos de variables son las formadas por las imágenes estructurales y las etiquetas de estas imágenes. Una vez que se establece la transformación, la matriz de "scores" estará formada por las proyecciones de los voxels que se han obtenido teniendo en cuenta las etiquetas de las imágenes de diagnóstico. Estas proyecciones, llamadas componentes PLS, se utilizan como vectores de características para la etapa de clasificación.

El segundo sistema CAD propuesto para la detección de la EA se basa en el análisis de componentes principales (PCA, del inglés "Principal Component Analysis"), como técnica de extracción de características para las imágenes cerebrales de sMRI. Este enfoque reduce la alta dimensión del espacio original de las imágenes cerebrales mediante una transformación a un subespacio de menor dimensión. PCA genera un conjunto de vectores de base ortonormal, conocidos como componentes principales, que maximiza la dispersión de todas las proyecciones de las muestras, lo que equivale a encontrar los valores propios de la matriz de covarianza.

El tercer sistema se basa en la técnica del análisis de componentes independientes (ICA, del inglés "Independent component analysis"). En primer lugar, definimos una "plantilla" de referencia para el sistema CAD con objeto de seleccionar las regiones de interés. Estas plantillas fueron definidas siguiendo dos aproximaciones:

la plantilla promedio de sujetos normales y la plantilla diferencia entre sujetos normales y patológicos. Este enfoque se utiliza para extraer los componentes independientes que revelan los patrones de variación del conjunto de datos.
El último sistema CAD se basa en la técnica de descomposición conocido como factorización de matrices no negativas (NNMF, del inglés "Non-Negative Matrix Factorization") que extrae información significativa mediante un proceso de optimización (minimización) de una función de error con el objectivo de reducir la dimensión. Este enfoque representa el conjunto de datos no negativos como una transformación lineal de variables con valores positivos. Después de la etapa de factorización, los datos transformados tienen un rango inferior a los datos originales. Por tanto, un conjunto pequeño de variables representa los datos de cada perfil en el nuevo espacio de características. Estas técnicas de extracción de características resuelven con éxito el problema del pequeño tamaño muestral mediante la extracción de la información relevante relacionada con la enfermedad. Este proceso se conoce como la reducción de la dimensionalidad y mejora la precisión de la predicción de los sistemas de diagnóstico asistido por ordenador, especialmente en la etapa temprana de la enfermedad. Finalmente, los sistemas propouestos fueron validados mediante una técnica de remuestreo (en inglés, K-fold cross-validation) con el fin de ajustar los parámetros del clasificador y estimar su rendimiento. Teniendo en cuenta el hecho de que un diagnóstico precoz de la demencia es crucial, los experimentos de clasificación se realizaron no sólo para distinguir sujetas normales y sujetos con EA, sino también para diferenciar NC de una fase de transición entre esta clase y la de los sujetos que padecen EA. Esta última clase se denomina deterioro cognitivo leve (MCI, del inglés "Mild Cognitive Impairment"). Los resultados obtenidos demuestran la efectividad y la robustez de los sistemas CAD propuestos (con una precisión superior al 87 %) en comparación con técnicas anteriores como el método basado en los vóxeles como características (VAF, del inglés, Voxels-As-Features).

Abbreviation

AD: Alzheimer's Disease

ADAS-Cog: Alzheimer's Disease Assessment Scale–Cognitive subscale

ADNI: Alzheimer's Disease Neuroimaging Initiative

CAD: Computer Aided Diagnosis

CAD: Computer Aided Diagnosis

CDR: Clinical Dementia Rating Scale

CSF: Cerebrospinal Fluid

CV: Cross Validation

FP: False Positives

FN: False Negatives

FPP: False Prediction of Positive

FPN: False Prediction of Negative

GDS: Global Dementia Scale

GM: Grey Matter

ICA: Independent Component Analysis

ICAm: Independent Component Analysis on means

KL: Kullback-Leibler

LSVM: Linear Support Vector Machine

LOO: Leave-One-Out

MCI: Mild Cognitive Impairment

MRI: Magnetic Resonance Imaging

MMSE: Mini Mental State Examination

MNI: Montreal Neurological Institute

MWW: Mann-Whitney-Wilcoxon

NMR: Nuclear Magnetic Resonance

NNMF: Non-Negative Matrix Factorization

NPV: Negative Predictive Value

NC: Control Subject

PCA: Principal Component Analysis

PLS: Partial Least Squares

PPV: Positive Predictive Value

ROC: Receiver Operating Characteristic

ROI: Region of Interest

RBF: Radial Basis Function

sMRI: structural Magnetic Resonance Imaging

SNR: Signal-To-Noise Ratio

SPM: Statistical Parametric Mapping

SPM8: Statistical Parametric Mapping Software, version 8

SVM: Support Vector Machine

TN: True Negatives

TP: True Positives

VAF: Voxels-as-Features

VBM: Voxel Based Morphometry

WM: White Matter

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this introductory chapter, we want to show, on the one hand, the need for developing automatic systems to aid the diagnosis of neurodegenerative dementia as a principal motivation in this PhD work. In this sense, we present a brief description of the state of the art of the disease under study as well as the computer aided diagnosis (CAD) systems which help the specialist to establish an early diagnosis of the disease. Developed CAD systems based on the selection of the region of interest in structural tomography imaging will be presented in detail in the following chapters. On the other hand, the objectives and the contributions of this thesis are cited in the following sections of this chapter, as well as, the organization of this PhD report.

1.1 General Introduction

Dementia is one of the most common neurodegenerative disorders among elderly people and has dramatic health consequences and major socio-economic implications [1–3]. Due to the aging of the population in the developed countries, the relevance of this disease increases every day [4, 5]. Statistical research demonstrates that the number of patients is expected to double in 2020 and triple by 2050 [6, 7]. In recent years, the use of computer systems based on tomographic brain images has made significant improvements in the diagnosis of the dementia. However, the processing of tomographic images still has a large margin for improvement.
Recent research suggests that the neurodegenerative process associated with dementia begins several years before patients have symptoms [9]. In the last decade, the use of magnetic resonance imaging (MRI) is generalized for diagnosing different types of dementia. Typically, these images are evaluated by experts who visually assess the presence of the typical characteristics of the disease. However, visual examination of the images remains a subjective process requiring experienced clinicians. For this reason, the development of CAD systems is one of the major research subjects to help doctors in the analysis of tomographic medical imaging, specifically, in the early detection of neurodegenerative disease [10–12]. CAD system [13] is performed through automated analysis techniques. It helps clinicians to take the diagnostic decisions. Hence, CAD system provides additional information for an accurate diagnosis [14–16] and eliminates the subjectivity inherent in the exploration of visual images. In this sense, two widely used methods for the development of a CAD system have been used in the literature for the analysis of neuroimaging data. The first one is based on univariate statistical testing. Therefore, this later is performed on Statistical Parametric Mapping (SPM) which compares the voxel values of the image under study to the mean values of the normal subjects. However, its application to distinguish between different groups reports poor classification results. The second one is based on the multivariate approach which considers all voxels of the brain image as a single observation. Furthermore, most of these multivariate approaches use only a small set of voxels (or reduced regions) to distinguish between pathological and normal control images [17, 18]. The classification step of multivariate approach is done by defining feature vectors representing the different images and training a classifier with a given set of known samples [19–21]. This step is used to distinguish between normal control subjects and Alzheimer patients. SVM classifier has been widely used successively in the pattern recognition field such as in the medical imaging diagnosis application [21, 22]. The application of the multivariate approach reports a high classification accuracy. Recently, this approach is widely used due to recent advances in defining the statistical classifiers to build more reliable and more capable of generalization classifiers, able to better address the problem of small sample size [23].
One of the simplest multivariate approaches for developing a CAD system for AD

is the well-known Voxel-As Features (VAF) method [19], in which all voxels are considered as features [19, 20], and then used as input to the classifier. However, many more advanced CAD systems can be made by adding different combinations of feature selection and feature extraction algorithms. In this way, the research group of Signal Processing and Biomedical Applications (SIPBA) of the University of Granada has made important advances in the study of early detection for neurological disorders by developing preprocessing techniques, analysis and classification of several types of biomedical imaging. Several optimization methods based on Levenberg-Marquardt and Gauss-Newton models for spatial normalization of functional brain images were tested in [24, 25]. In addition, novel voxel selection models using the standard deviation and the Mann-Wilcoxon techniques were proposed, as well as a method for calculating the regions of interest using the factor analysis approach. Therefore, the work described in [70] optimized several intensity normalization methods for neuroimaging data based on Gaussian Mixture Model, mean square error and multivariate linear regression models. In [26], the PhD work deepens into the Principal Component Analysis (PCA) and the Independent Component Analysis (ICA) as feature extraction approaches. On the other hand, the candidate developed the PCA approach by the introduction of the Karhunen-Lo'eve transformation and the selection of eigenbrains by the criterion of Fisher. In addition, it delves into the development of learning methods such as the SVM classifier. Furthermore, a clustering method based on Gaussian mixture model (GMM) is proposed in [14]. The goal of this work is to extract the significant voxels that define the regions of interest (ROI) for further classification. In addition, a partial least squares (PLS) approach was introduced in the above reference to extract the relevant features from functional PET and SPECT imaging. Thus, a significant improvement has been obtained for the classification.

The main contribution of this thesis is the development of new strategies to overcome the small sample size problem in neuroimaging. This idea can improve the accuracy rate of the CAD systems by reducing the number of false positives, increasing the reliability of their results. On the one hand, the feature space reduction is performed through the segmentation process in order to take into account the structural change in gray matter (GM) and white matter (WM) brain tissues, during the early stage of AD. On the other hand, the developed CAD systems are based on the development and the evaluation of feature selection and feature extraction approaches to extract only the relevant information related to AD. Thus, the main goal of the PhD thesis is to improve the prediction accuracy of CAD systems, specifically in the early stage of the disease. The pipeline of the developed CAD systems is presented in figure 1.1. These CAD systems consist of three different stages: preprocessing (spatial normalization, segmentation), post-processing (feature selection and feature extraction) and validation. After all, the MRI raw images are blurred and affected by several sources of artifacts and noise [27]. Thus, in order to correct non-uniformities of the magnetic field, the source images need to be preprocessed.

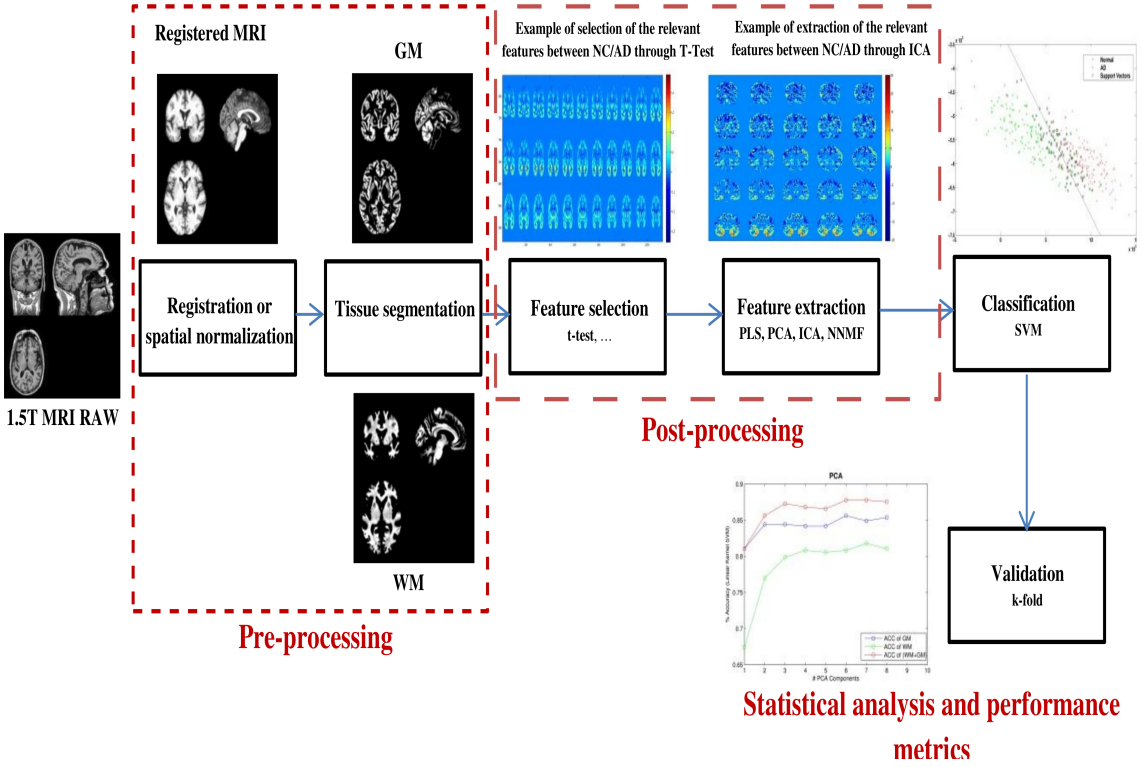


Figure 1.1: Flow chart of the classification process.

Moreover, in order to reduce the variability between subjects and allow meaningful group analyses on a voxel by voxel basis, the spatial normalization is performed. Afterwards, the segmentation process is realized for analyzing the change in each brain tissue. Feature selection and feature extraction tasks identify the relevant features that are used by pattern-recognition and machine-learning techniques for disease identification and validation. In this sense, the classification step is implemented using SVM to identify each subject image and the validation of the CAD system was estimated through a k-fold cross-validation strategy.

1.2 Motivation

MRI is a medical imaging process-based on the phenomenon known as Nuclear Magnetic Resonance (NMR), which is gaining widespread acceptance for a large variety of medical exploration [28]. It is one of the most commonly used techniques in radiology to visualize the body's soft tissues with great contrast, including the brain. Some reasons for this growing interest are:

- It is a non-invasive, based on non-ionizing radiation, imaging procedure.

- The fast evolution of MRI imaging technique offers a wide list of pulse sequences that can easily be tuned to offer specific visualizations. MRI has become a very flexible imaging tool.

- It has a high spatial resolution and provides a great deal of information on the anatomical structure, allowing quantitative pathological or clinical studies.

More recently, a variety of MRI imaging modalities including structural and functional MRI have shown characteristic changes in the brains of patients with AD, and in prodromal and even presymptomatic states that can help rule-in the AD pathophysiological process [29]. However, this work is focused only on sMRI. The latter has a high spatial resolution and provides a large amount of information on the anatomical structure.

This PhD thesis work is structured in two main parts. In the first part, the background of the work is introduced, as well as the pattern recognition technologies that are based on the contributions in the field of statistical learning. Thus, we start by describing the pathology under study, the Alzheimer's disease, the effects of this disease in our society and the importance of the early diagnosis. Then, the tomography techniques are described, that provide us the "maps" of brain activation; keys to diagnosis it with computer system known as the CAD systems. Following this, we describe the state of the art in the field of neuroscience for the automatic diagnosis and/or the quantitative evaluation of brain images. Finally, the databases on which we will work are described, as well as the pre-processing step that is used to gain information from them.
In the second part, four developed CAD systems are proposed and described to improve the diagnosis of the disease, based on partial least squares (PLS), principal components (PCA), independent component (ICA) and Non-negative matrix factorization (NNMF) approaches. Moreover, the obtained experiment results are presented in detail in this part of the thesis.

1.3 Goals

The main objectives of this work are aimed at the development of novel processing and classification techniques for structural brain magnetic resonance imaging. This leads to the construction of CAD systems to support decision making in the clinical context of neurodegenerative diseases diagnosis with particular attention to the Alzheimer's disease. Furthermore, the specific objectives are summarized in the development of advanced feature extraction techniques and automatic classification of sMRI to identify patients affected by Alzheimer's disease. The design and

the evaluation of CAD systems are carried out by using the ADNI (Alzheimer's disease Neuroimaging Initiative) database, a multicenter study aimed to evaluate the biomarkers that describe the progress of the AD. Thus, it provides sMRI biomarkers of control subjects, patients with cognitive impairment (MCI) and AD patients for detecting the disease in its early stage.

In this PhD work, new efficient processing and classification techniques of sMRI are developed with double objectives to:

a) Improve the sensitivity value for AD pattern detection, specifically, in its early stage through the development of systems that are based on previous image labeling performed by clinical experts.

b) Reduce the computational time of the diagnostic by the implementing of efficient algorithms that overcome the limitations of the high dimensionality of neuroimaging data.

1.4 Main contributions

The main scientific contributions of the thesis can be split into the development of new CAD systems for early detection of AD and the improvement of the accuracy rate. These CAD systems are based on standard approaches for selecting the region of interest related to the disease. In the following section, the major scientific contributions are briefly introduced. More details are given in the later chapters of this thesis. The areas covered are:

- The description of the state of the art of the Alzheimer's disease.
- The preprocessing steps to analyze the sMRI data (spatial normalization, smoothing, segmentation).
- The application of feature selection and feature extraction techniques to the segmented sMRI data.
- The application of classification approaches for early detection of AD based on sMRI brain images.
- The validation of these developed CAD systems through k-fold cross validation methodology.

We have proposed four CAD systems based on different feature extraction and selection techniques in order to support the clinical diagnosis of Alzheimer's disease.

- The first and the second proposed CAD systems are based on t-test feature selection technique for selecting the region of interest, and on the PLS and the PCA feature extraction techniques respectively to solve the small sample size problem. The extracted features are combined with supervised classification methods based on SVM to classify the segmented sMRI database.

- The third CAD system is based on two ICA models that are proposed to extract the relevant features related to AD. Thus, the first model is based on the extraction of a low number of Independent Components (IC) which work as feature vectors for each brain tissue image. The second one is based on the extraction of highly representative features from each average brain image (NC, MCI and AD) from each brain tissue. Then, the set of independent component sources are used as input variables for the classification step of early detection of the disease.

- The fourth CAD system is based on NNMF approach for finding reduced linear representations of non-negative data. Then, the resulting NNMF-transformed sets of data, which contain the reduced relevant features, are classified by means of different SVM-based classifiers.

1.5 Publications produced during the PhD thesis

Part of the work presented here has been published and is already available for the research community.

Articles in International Magazines:

1. L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim and F. Segovia. Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. Journal of *Neurocomputing*, 151 (1): 139–150, 2014, DOI:10.1016/j.neucom.2014.09.072.

2. L. Khedher, I. A. Illán, J. M. Górriz, J. Ramírez, A. Brahim and Anke Meyer-Baese. Independent Component Analysis-Support Vector Machine-based Computer - Aided Diagnosis System for Alzheimer's with visual support. Journal of *Neural Systems*, 27 (3): 1–19, 2016, DOI: 10.1142/S0129065716500507.

International Conferences Proceedings:

1. L. Khedher, J. Ramírez, J. M. Górriz , A. Brahim. Automatic classification of segmented MRI data combining Independent Component Analysis and Support Vector Machines. *International Conference on Innovation in Medicine and Healthcare (Inmed14)*, San Sebastian, Spain, 207: 271–279, 2014, ISBN: 978-1-61499-473-2.

2. L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim and I.A. Illán. Independent Component Analysis-Based Classification of Alzheimer's Disease from Segmented MRI Data. 6th. *International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC)*, Elche, Spain, 9107: 78–87, 2015, ISBN: 978-3-319-18913-0.

Several papers were published with collaborations in different international magazine and conference proceedings:

1. A. Brahim, J. Ramírez , J. M. Górriz, L. Khedher and D. Salas-Gonzalez. Comparison between Different Intensity Normalization Methods in ^{123}I-Ioflupane Imaging for the Automatic Detection of Parkinsonism. Journal of *Plos One*, 10 (6): 1–20, 2015, DOI:10.1371/journal.pone.0130274.

2. A. Brahim, J. M. Górriz, J. Ramírez, L. Khedher. Intensity normalization of DaTSCAN SPECT imaging using a model-based clustering approach. Journal of *Applied Soft Computing*, 37: 234–244, 2015, DOI/10.1016/j.asoc.2015.08.030.

3. A. Brahim, J. M. Górriz, J. Ramírez, L. Khedher. Linear intensity normalization of DaTSCAN images using Mean Square Error and a model-based clustering approach. *International Conference on Innovation in Medicine and Healthcare (Inmed14)*, San Sebastian, Spain, 207: 251–260, 2014, ISBN: 978-1-61499-473-2.

4. A. Brahim, J. Ramírez , J. M. Górriz, L. Khedher. Linear intensity normalization of DaTSCAN images using Mean Square Error and a model-based clustering approach. *IEEE International Conference on Image Processing (ICIP14)*, Paris, France, 207: 3617–3621, 2014, ISBN: 978-1-4799-5751-4.

5. A. Brahim, J. M. Górriz, J. Ramírez, L. Khedher. Intensity Normalization of ^{123}I-ioflupane-SPECT Brain Images Using a Model-Based Multivariate Linear Regression Approach. *6th. International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC)*, Elche, Spain, 9107: 68–77, 2015, ISBN: 978-3-319-18913-0.

1.6 Research project

During this thesis, the candidate has been part of several research projects which are related to the contents of the PhD thesis.

1. Project Title: "Plataforma Abierta de Procesamiento de Imágenes para ayuda al diagnóstico de Alteraciones Neurológicas (PAPI-ADAN)". Excellence Regional Project: P09-TIC-4530, 2012–2014. Ministry of Economy, Innovation, Science and Employment (Spain).

1. Project Title: "Modelos estadísticos de neurodegeneracion para sistemas de ayuda al diagnóstico (STM- NEUROCAD). Aplicación al diagnóstico precoz de las enfermedades de Alzheimer y Parkinson". Excellence Regional Project: P11–TIC–7103, 2012–2016. Ministry of Economy, Innovation, Science and Employment (Spain).

1.7 Structure of the document

Following the introduction, the PhD report is divided into three main parts. The first part contains five chapters that present the medical imaging and its use for the diagnosis of neurodegenerative diseases, such as AD. Moreover, the different techniques of diagnosis and classification are presented:

- Description of the Alzheimer disease.

- Brief presentation of the tomographic techniques used in nuclear medicine, focusing on structural imaging techniques.

- Details about the structural image database used in this work in order to validate the proposed CAD systems.

- Description of the diagnostic criteria and the computerized techniques that have been developed to date in order to obtain a CAD systems for AD.

- Classification and performance assessment: description of classification methods and validation systems used in this work.

The second part includes five chapters:

- Chapter 6, chapter 7, chapter 8 and 9 present the developed CAD systems proposed for automatic detection of Alzheimer's disease. These systems based on reduction dimensionality and compression techniques to extract the relevant features for the classification task.

- Finally, the performance of the developed image-based CAD systems are discussed in chapter 10. Moreover, several conclusions and possible paths for future research are presented.

The third and the last part is a summary in Spanish of the presented PhD work.

Part I

Fundamental theory

Chapter 2

Alzheimer's disease

Dementia, one of the most severe and frequent neurodegenerative disorders in the elderly population, has important and dramatic health as well as socio-economic implications [1]. Recently, Alzheimer's disease (AD) is considered as one of the most common cause of dementia. In this sense, the development of new methods for analyzing tomographic images can improve the early diagnosis of AD. This chapter provides an overview of this disease, its causes and its neuro-pathology evolution. In addition, the different traditional diagnostic techniques used for the diagnosis of the disease are presented in the end of the current chapter.

2.1 Alzheimer's disease

AD is the most common cause of dementia in the elderly and is characterized by a spectrum of clinical features and neuropathological assessment that appear progressively [30]. At the beginning of the 20th century, the German neurologist Max Bielchowsky visualizes cellular components inside of the neurons that are called neurofibrillary tangles. In 1906, using the Bielchowsky method, Dr. Alois Alzheimer described the beginning of a new pathology in the brain of one of her patients: Auguste Deter, 51 years [31]. Series of clinical symptoms are presented in the patient, including weakness of the memory skills, loss of orientation, deep agitation, a pronounced psychosocial disability [39]. The rapidity of degeneration and the relative youth of the patient motivate the Dr. Alzheimer to investigate her neuropathological characteristics [32]. Hence, he found many abnormal clumps (now called amyloid plaques) and tangled bundles of fibers (now called neurofibrillary neurofibrillary tangles). These plaques and tangles in the brain are nowadays considered signs of AD [32]. AD is a slow disease, starting with mild memory problems and ending with severe brain damage. The evaluation and the progression of the disease vary from person to person, although, Alzheimer's patients live between 8 and 10 years after the disease is detected [33].

Nowadays, AD is assumed to be the most common cause of dementia in the elderly. Besides, the World Health Organization estimated in 2005 that 0.379% of people worldwide had dementia, and that the prevalence would increase to 0.441% in 2015 and to 0.556% in 2030 [7, 34].
Figure 2.1 indicates an other recent statistical number of people will be affected by AD in 2050. As shown in this figure, the number of people affected by the disease will continue growing in the next decades. For this reason, it is important to better understand the disease, to really understand the causes, and to develop treatments that can stop the progression of the disease or even cure it.

2.1.1 Causes

Until now there are no consensus in the scientific community about the causes or the risk factors of Alzheimer's disease. In some cases (less than 2 %), the disease is caused by genetic mutations in the family. In these cases, the symptoms begin before the age of 60 and it progress rapidly [35]. All the known mutations involve an overproduction of a protein that destroys the nerve cells. Also, it is known that some genetic factors increase the chances that a person will develop the disease. Specifically, the carriers of a variant of the ApoE gene, which promotes the creation of preliminary proteins. This latter is more susceptible to developing the disease in the brain [36, 37].

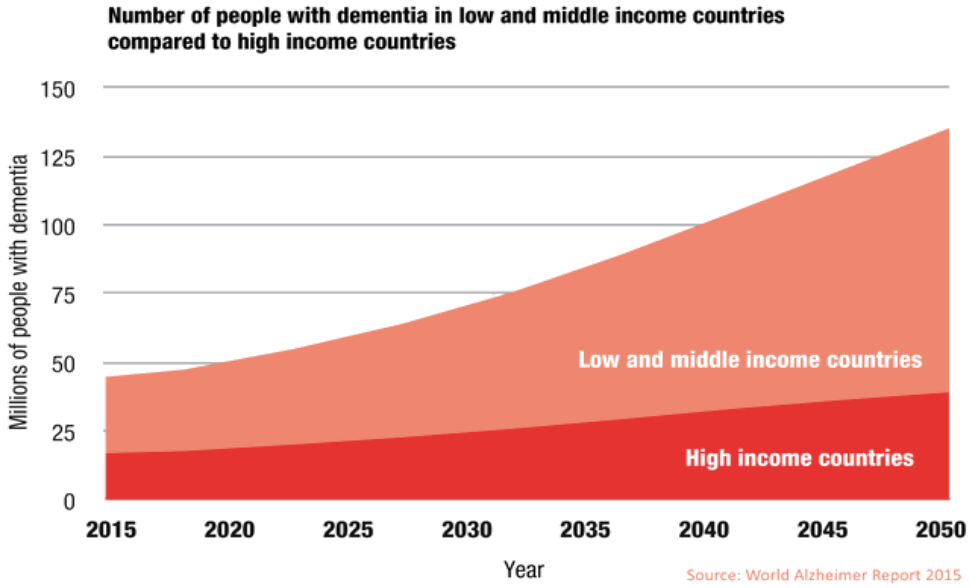


Figure 2.1: Estimated number of people with dementia until 2050 in high, middle and low income countries[8].

On the one hand, age is the most important risk factor. The number of people over 65 years with the disease doubles every 5 years [38]. On the other hand, people with less education are more likely to suffer from this dementia. Other factors such as environment or diet seem to have no influence, however, research studies still be continued.

2.1.2 Neuro-pathology evolution

Until now, the real causes for AD still remain unclear, some is known about the evolution of the pathology and its physical influence on the brain. The basis of the pathology of this disease includes neuronal degeneration and the formation of lesions in the brain. Figure 2.2 shows an example of brain pathology of subject with AD and subject without dementia.

2.1.2.1 Causes

AD is mainly characterized by the appearance of lesions in the brain. These lesions, called amyloid plaques and neurofibrillary tangles, appear in some regions and it gradually expand over most of the cortex. The formation of these lesions are thought to contribute to the degradation of the neurons in the brain and the subsequent symptoms of Alzheimer's disease (see figure 2.3).

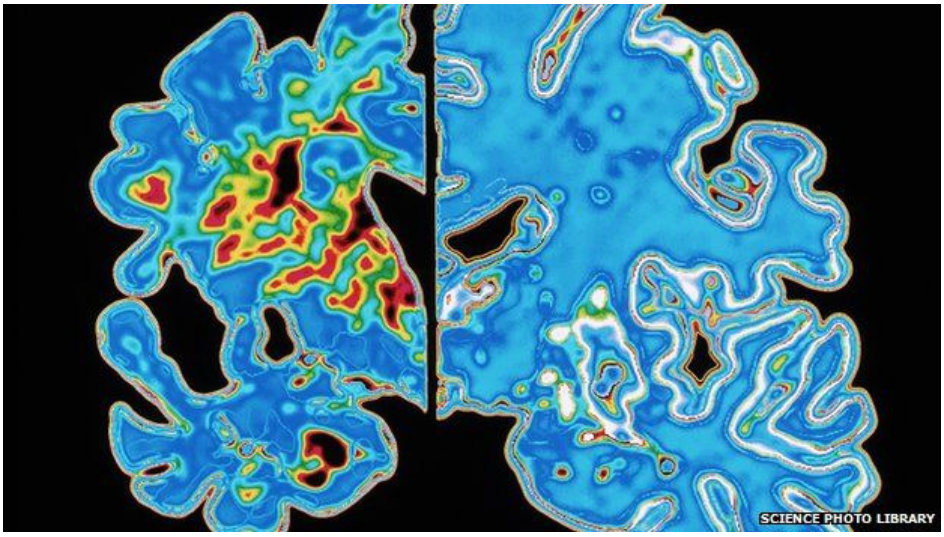


Figure 2.2: Structural differences between a brain affected by AD (left) and healthy (right) [39].

The two major pathological features of AD are extracellular plaques and intracellular tangles, accumulated between neurons in the brain. Amyloid is a general term for protein fragments that the body produces normally. In a healthy brain, these

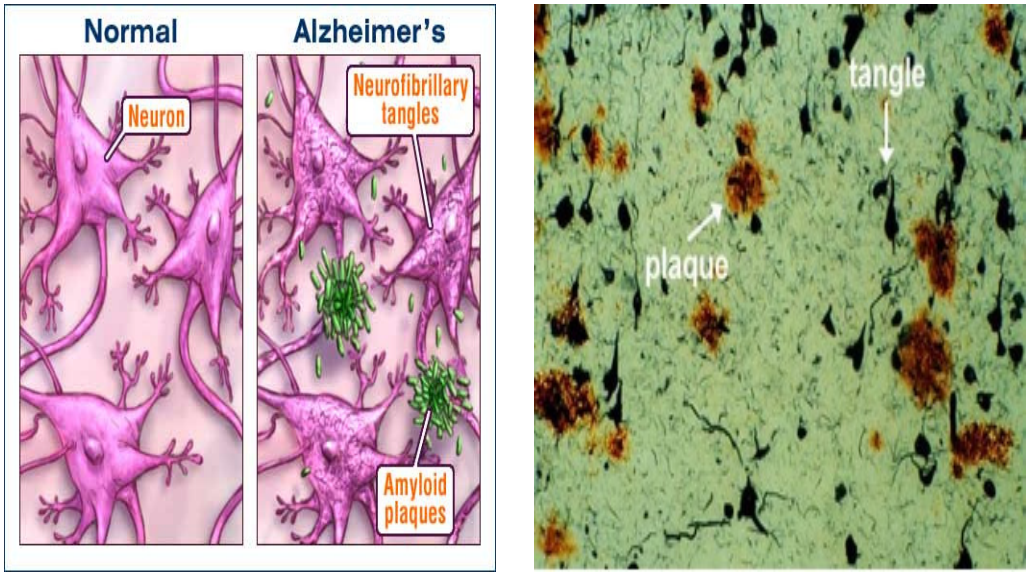


Figure 2.3: Alzheimer's disease neuropathology: Degeneration changes in Alzheimer's diseased brain. Apparition of amyloid plaques and neurofibrillary tangles (left). Microscopic view in the cerebral cortex of an Alzheimer's patient (right) [39].

protein fragments would be broken down and eliminated. In Alzheimer's disease brain, the fragments accumulate to form insoluble plaques. Neurofibrillary tangles consist of insoluble twisted fibers that are found inside of the neurons. They primary consist of a protein called tau, which forms part of a structure called a micro-tubule. The micro-tubule helps transport nutrients and other important substances from one part of the neuron to another. In a patient with Alzheimer's disease, the tau protein is abnormal and the micro-tubule collapses.

2.2 Mild cognitive impairment

Many patients with memory impairment do not necessarily meet the clinical criteria for dementia. The concept of mild cognitive impairment (MCI) is thought to be a transitional phase between being cognitively normal and having an AD diagnosis. Patients with MCI have a higher risk of developing AD than elderly with normal cognitive function [40]. The concept of MCI is rather difficult to describe since it is a very heterogeneous group. The diagnostic accuracy of the criteria today is low to moderate [41]. For this reason it has been divided into different categories, namely amnestic (having memory deficits) and non-amnestic (having no memory deficits, but other cognitive problems). These categories can be further split into single and multiple domains [42]. Single domain means that the subject has one cognitive deficit and multiple domains involve several different cognitive deficits.
The common criterion for describing a subject with MCI states that the subject has memory complaints and objective memory impairment but with overall preservation of the cognitive functions and handles daily life [43]. Thus, it can be difficult to diagnose accurately since it may be mistaken for normal aging [44]. 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 [9].
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. Besides, the clinical group of MCI is very important for both preventive trials and evaluating MR markers for early diagnosis and monitoring of disease progression [42, 45].

2.3 Clinical diagnostic techniques

Early diagnosis of this disease is crucial to improve and extend the life of the patient. Moreover, the early diagnosis can offer the best chance to treat the symptoms of the disease.
Currently, the only definitive way to diagnose the Alzheimer's disease is to re-

search the existence of the plaques and the tangles in the brain tissue. Therefore, the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's disease and related Disorders Association (ADRDA) outlined the clinical criteria of AD [46]. Thus, the NINCDS–ADRDA provides three levels of certainty about the diagnosis of AD: 'possible', 'probable' or 'definite'.
The criteria for possible AD is based on the observation of clinical symptoms and the deterioration of some cognitive functions (e.g. memory, language, etc). For the diagnosis of probable AD, dementia has been established by neuropsychological examination. Progressive cognitive impairment has to be present in two or more areas of cognition. The onset of the deficits has occurred between the ages of 40–90 years. There must be an absence of other diseases capable of producing dementia. Finally, to make the diagnosis of certain AD (or 'definite'), the identification of plaques and tangles in the brain confirm that the brain is affected by Alzheimer's disease dementia.
The diagnosis of dementia is mainly based on the clinical evaluation that requires a comprehensive assessment of cognitive function, specifically the memory, attention, perception, language, praxis. In addition, several other neuropsychological tools used to support in the diagnostic procedure. The neuropsychological assessment can be subdivided into two levels of complexity:

- The first level consists of the use of short, standardized and simple tests as the Mini-Mental State Examination (MMSE) [47], which enables the diagnosis of dementia.

- The second level of greater complexity in the assessment of the severity of the impairment is refined, while the cognitive function domains that are affected are established [48]. There are different scales that provide a standardized according to the degree of functional impairment, as CDR (Clinical Dementia Rating) [49] or GDS (Global Deterioration Scale) value [50].

In general, these scales allow to classify the dementia according to the classic clinical criteria: mild, moderate or severe dementia. The questionnaires or the scales have been designed to quantify certain cognitive functions, i.e, it do not establish a diagnosis but it quantify the severity of the alteration of certain intellectual areas, it is particularly valuable to discriminate between normal aging and mild dementia.
The diagnosis must always based on the clinical history of the patient and according to the criteria established. In this sense, the questionnaires represent only an aid in the process of the valuation of the disease.

2.3.1 Mini mental state examination

The Mini Mental State Examination (MMSE) [47] is one of the most commonly used tests for complaints of memory problems. The MMSE test is a 30-point questionnaire used for measuring severity and decline in cognition. It deals with function, such as memory function, calculation, language abilities and attention. Gaining a score of 24 –30 is considered as no dementia, 19 –23 as mild dementia and 13 –18 indicates moderate dementia.
In the dementia by the Alzheimer's disease, the average rate of change in the score of the MMSE is of 2-5 points per year, reason why the test shows its utility for the pursuit of the affected patients. The MMSE has a limited ability to detect non-Alzheimer's dementias, such as post-stroke cognitive impairment, frontotemporal or subcortical dementias in their early phases.
The essential characteristics which are evaluated as follows:

- Capacity of attention, concentration and memory.
- Capacity of calculation.
- Capacity of language and visoespacial perception.
- Direction space-time.
- Capacity to follow basic instructions.

This test provides an instrument for detecting cognitive impairment and can be performed in a short time. According to the clinical experts, this is a specially important in the diagnosis of the dementia.

2.3.2 Global dementia scale

The Global Deterioration Scale (GDS) [51], developed by Dr. Barry Reisberg, provides an overview of the stages of cognitive function for those suffering from a primary degenerative dementia such as AD. It is broken down into 7 different stages as shown in table 2.1.

Table 2.1: The different stages which establishes the global dementia scale [52].

Level	Cognitive deficit	Clinical Characteristics
1	No cognitive decline	No subjective complaints of memory deficit. No memory deficit evident on clinical interview.
2	Very mild cognitive decline	No objective evidence of memory deficit on clinical interview. No objective deficits in employment or social situations. Appropriate concern with respect to symptomatology.
3	Mild cognitive decline	Objective evidence of memory deficit obtained only with an intensive interview. Decreased performance in demanding employment and social settings. Denial begins to become manifest in patient. Mild to moderate anxiety accompanies symptoms.
4	Moderate cognitive decline	Inability to perform complex tasks. Denial is dominant defense mechanism. Flattening of affect and withdrawal from challenging situations frequently occur.
5	Moderately severe cognitive decline	Persons at this stage retain knowledge of many major facts regarding themselves and others. They invariably know their own names and generally know their spouses' and children's names. They require no assistance with toileting and eating, but may have some difficulty choosing the proper clothing to wear.
6	Severe cognitive decline	May occasionally forget the name of the spouse upon whom they are entirely dependent for survival. Will be largely unaware of all recent events and experiences in their lives. Retain some knowledge of their past lives but this is very sketchy.
7	Very severe cognitive decline	All verbal abilities are lost over the course of this stage. Frequently there is no speech at all -only unintelligible utterances and rare emergence of seemingly forgotten words and phrases. Incontinent of urine, requires assistance toileting and feeding.

2.3.3 Clinical dementia rating

The Clinical dementia rating (CDR) [44] scale is also used for the evaluation of staging severity of dementia. This is a five point scale, rating six functional domains. These are memory, orientation, judgment and problem solving, community affairs, home and hobbies and finally personal care. The five point scale is as follows:

– CDR=0: no dementia

– CDR=0.5: very mild dementia

– CDR=1: mild dementia

– CDR=2: moderate dementia

– CDR=3: severe dementia

2.3.4 Alzheimer's disease assessment scale

The Alzheimer's Disease Assessment Scale (ADAS) is a two part scale designed to assess cognitive and non-cognitive symptoms of AD. It is one of the most frequently used scales in clinical trials but is quite time-consuming (taking on average 45 minutes to complete). The part which measures cognitive faculties is known as the ADAS-Cog [50]. It is a subscale of ADAS, which was designed to measure the severity of the most important symptoms of AD. This subscale is a popular cognitive testing instrument used in clinical trials. This test includes 11 tasks measuring cognitive disturbances, language, attention and other cognitive abilities. Finally, brain scan examinations are performed on routine bases to aid in the clinical evaluation. MRI is today the first choice in many countries.

2.4 Conclusion

Neurodegenerative diseases are becoming increasingly prevalent with the aging of general population. Alzheimer's disease (AD) is most prevalent of neurodegenerative disease. An overview of this disease is presented in this chapter. Moreover, the different traditional diagnostic techniques for AD is introduced in detail in this chapter.

Chapter 3

Functional/Structural Imaging modalities

Neuroimaging techniques provide a way for clinicians to examine the structural and functional changes in the brain associated with the development of Alzheimer's disease (AD). Many imaging modalities are commonly used, such as, magnetic resonance imaging (MRI), X-ray computed tomography (CT), positron emission tomography (PET), single-photon emission computed tomography (SPECT), and diffusion tensor imaging (DTI) (figure 3.1). In the framework of this thesis, we focus mainly on a structural MRI dataset which is described in this chapter.

3.1 Fundamentals

Neuroimaging techniques allow to see images of the central nervous system in general and of the brain in particular. These brain imaging techniques provide additional information for diagnosis and treatment of patients with disorders of the nervous system [53]. These techniques can be classified as functional or structural, according to functional brain information (ie. cerebral blood flow, glucose metabolism, amyloid deposition) [54] and structural brain information (ie. tissue and organ shape, texture) that offer [55]. The structural image, such as, structural magnetic resonance imaging (sMRI), provides a view of the brain structure to allow the diagnosis of diseases on a large scale. Furthermore, structural MRI provides information to describe the shape, size, and integrity of gray and white matter structures in the brain. On the other hand, functional neuroimaging techniques such as functional magnetic resonance imaging (fMRI), PET and SPECT have the ability to identify patho-physiological changes in the brain. These functional images are based on the measure of brain activity.
In neuroimaging, the most extended is by far MRI, which provides intensity maps that represent the internal structure of the brain. Particular attention has been paid in this work to the sMRI due to it is widely used in the clinical evaluation of patients with AD.

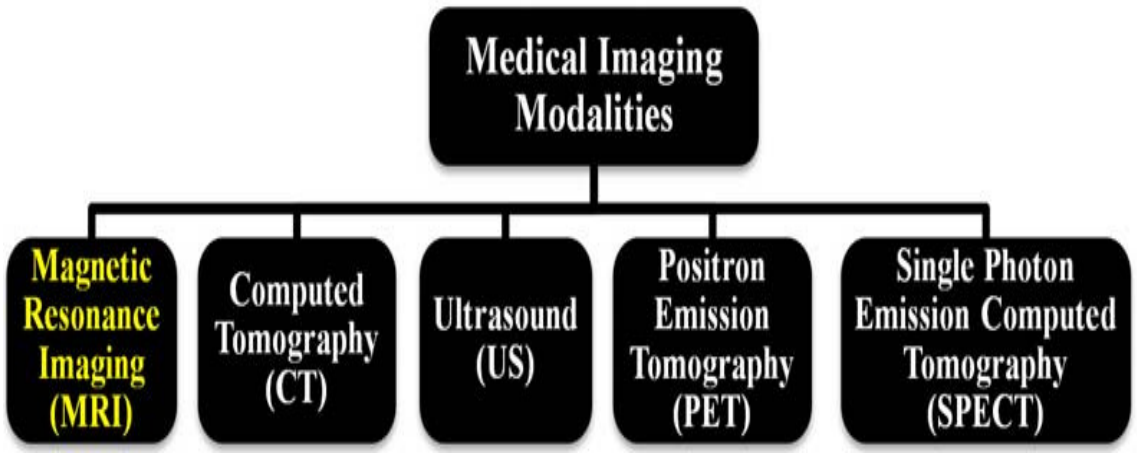


Figure 3.1: Schematic illustration of different classes of medical image modalities [56].

3.2 Magnetic resonance imaging

MRI is perhaps the most widespread imaging modality in neuroimaging [57], given its ability to visualize both structural and functional properties of the brain. It has become the most powerful and non-invasive tool for clinical diagnosis of diseases [57]. MRI uses strong magnetic fields to excite certain atomic nuclei, that can absorb and emit this energy. MRI combines a constant magnetic field with a radiofrequency (RF) emission to excite the atomic nuclei present in corporal structure, resulting in an image of the distribution of certain atoms in the body [58–60]. Most MRI systems use hydrogen atoms, since they are present in water (which adds up to around 70% of body mass) and the signal derived is stronger than other atoms, increasing the Signal-To-Noise Ratio (SNR) and processed to obtain a 3D grey-scale image. The procedure uses a strong magnetic field B_0 to align the magnetic moment of the hydrogen nuclei in parallel or anti-parallel (depending of their initial spin). In this sense, the magnetic moment of all nuclei will increase up to a stable state, in contrast to their null value in absence of the magnetic field B_0 . Within this magnetic field, the hydrogen atoms precess around an axis along the direction of the field. A given nuclei has a resonance frequency which is proportional to the intensity of B_0, allow us to resonate hydrogen far below potentially damaging frequencies. The precession frequency is known as the Larmor frequency in the MR community. The Larmor equation expresses a connection between the resonance frequency and the magnetic field, and it is said to be the most important equation in MR:

$$f = \gamma B_0 \tag{3.1}$$

The equation tells us that the frequency f is proportional to the magnetic field, B_0. The proportionality factor is 42 MHz/T for protons. It is called "the gyromagnetic ratio" or simply "gamma". When a subject is introduced in the MRI scanner, it is submitted to the magnetic field B_0, so that the hydrogen nuclei are aligned to the field, with a precession frequency f. Then, a RF pulse of the same frequency is generated, which is then absorbed by the nuclei, forcing them to place perpendicular to the field. Once the RF emission is interrupted, the nuclei return to its equilibrium state by means of a procedure called relaxation. In this procedure, they emit part of the absorbed energy, which is then captured by a RF receptor. Usually, position information is encoded in the RF signal by varying B_0 using gradient coils.

The RF signal is measured during the relaxation time, and two different relaxation times are set: the T1 (spin-lattice) relaxation time and the T2 (spin-spin) relaxation time. The T1 time is the time during which nuclei emit energy to the adjacent tissue and realign to the longitudinal plane (z axis), whereas the T2 time refers to the time when nuclei realign to the transversal plane (y axis). These times are used to create T1-weighted and T2-weighted images (see figure 3.2). T1-weighted images allow to

distinguish between grey matter (GM) and white matter (WM) in the cerebral cortex, to identify fatty tissue, and generally, obtain structural information. However, T2-weighed images are used to assess cerebro-spinal fluid (CSF) or to visualize and identify WM lessions.

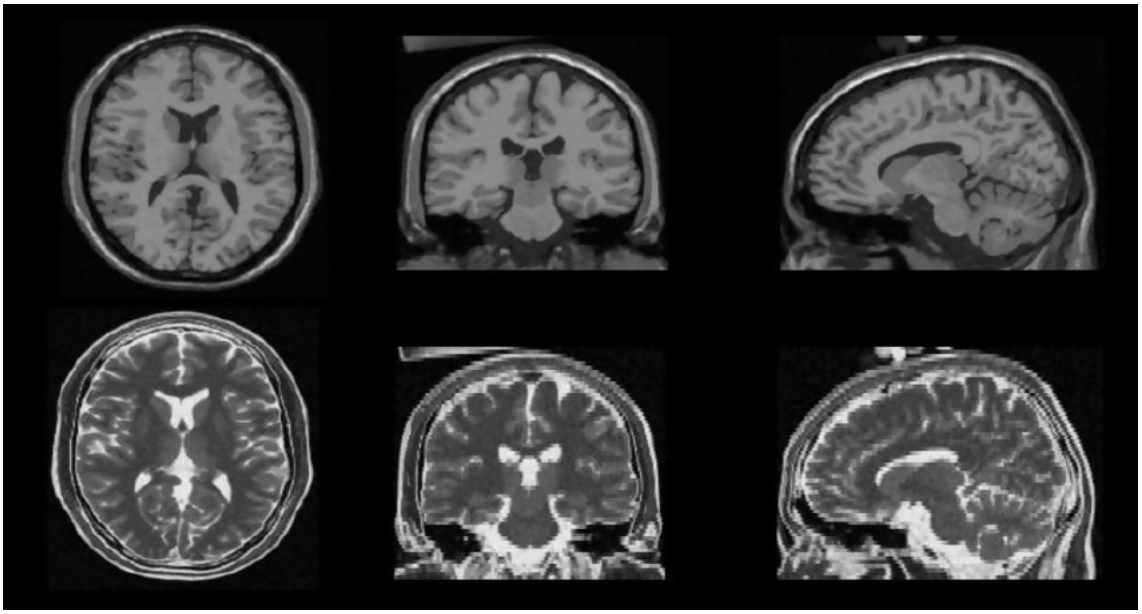


Figure 3.2: Example of MR data [61]. First row: high resolution T1-weighted image (1 mm slice thickness); second row: low resolution T2-weighted image (3 mm slice thickness).

3.2.1 Advantages

The MRI technique has several advantages compared to other neuroimaging techniques:

- It is fast and it does not use ionizing radiation.
- It can be repeated several times on the patients because the absorbed radiation is minimal.
- Its isotropic resolution is around $1mm^3$ with 3T MRI scanners, which outperforms the $8mm^3$ of PET.
- It has a high versatility, since it can be used to study structural and functional brain features with different configurations.

- It has the advantage of being able to visualize anatomy in all three planes: axial, sagital and coronal.

- It is not affected by the hardening beam effect of CT [62, 63] because the range of frequencies is small, and the attenuation coefficient of the tissues is almost homogeneous.

3.2.2 Disadvantages

- It is an expensive and complex technique.

- There are many parameters that must be tuned up correctly in order to optimize the image acquisition depending on the circumstances [63, 64].

- All the metal objects of the patients should be removed before the scanning starts, which is impossible for some kind of surgical implants.

- This technique is only suited to analyze soft tissues because the bones have not a significant contrast in the images.

3.3 The ADNI dataset

All the developed approaches in this PhD thesis have been validated by a structural MRI dataset that was acquired for the study of AD. This dataset is labeled by experts. The labels have been assigned to distinguish between images of AD patients (in different stages of the disease), images of people with mild cognitive impairment (MCI) that could lead to Alzheimer (it can be considered an early stage of disease) and images of people without the disease (NC; normal control subjects).
The structural MRI dataset was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. All details of the specific characteristics for this database are described in the following subsection.

3.3.1 Description of the database

The Alzheimer's Disease Neuroimaging Initiative (ADNI) was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and

clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and up to 1500 adults (ages 55 to 90) were recruited from over 50 sites across the U.S. and Canada in ADNI and its following initiatives ADNI-GO and ADNI-2 (see figure 3.3). Subjects had completed at least 6 years of education, and were fluent in Spanish or English. For up-to-date information on inclusion/exclusion criteria and other topics, see `www.adni-info.org`. In this the-

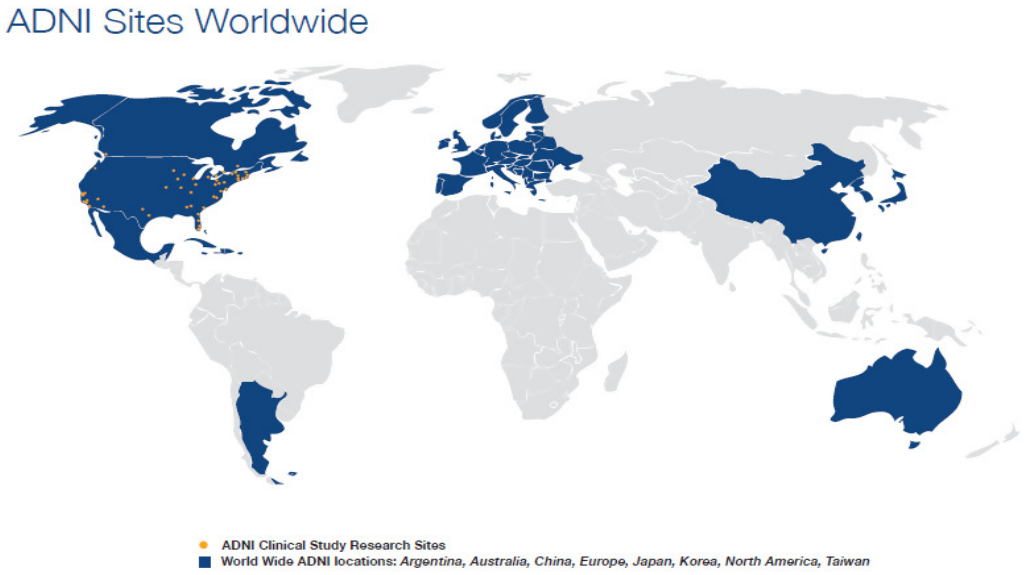


Figure 3.3: World wide Alzheimer's Disease Neuroimaging Initiative.

sis we will use data belonging to the ADNI-1 initiative. In particular, the database that we call adni-mri correspond to the MRI volumes from the 'ADNI1: Screening 1.5T' collection (subjects who have a screening data). It contains 818 T1-weighted MRI images from Normal Control subjects (229), mild cognitive impairment subjects (401; 312 stable MCI and 86 progressive MCI) and Alzheimer's disease subjects (188) (see demographic details at table 3.1). Information concerning the conversion from MCI to AD is taken from clinical data available from ADNI. Those patients whose clinical diagnosis suffer multiple conversions and reversions are considered as not reliably labeled and discarded from the MCI-converters cohort. Depending on the experiment, we may use spatially normalized T1-weighted MRI images, using the SPM8 software, and segmented GM and WM maps (see section 4.1). Segmentation was performed using the SPM-VBM8 toolbox .

Table 3.1: Demographic data of patients in the database (ADNI 1075-T1).

Diagnosis	Number	Age	Gender (M/F)	MMSE
NC	229	75.97±5.0	119/110	29.00±1.0
MCI	401	74.85±7.4	258/143	27.01±1.8
AD	188	75.36±7.5	99/89	23.28±2.0

3.3.2 Labelling criteria

The selection criteria followed for accepting participants in the ADNI project is based on a series of interviews and tests performed individually [46]. The results of the candidates had to meet certain conditions for admission to the project. Below are detailed criteria selection of patients for each of the classes of interest for the study:

1. Normal control subjects: Mini-Mental State Examination (MMSE) ([47]) scores between 24 and 30 (inclusive), a Clinical Dementia Rating (CDR) of 0, non-depressed, non-MCI, and non-demented. The age range of normal subjects was roughly matched to that of MCI and AD subjects. Therefore, there should be minimal enrollment of normals under the age of 70.

2. MCI subjects: MMSE scores between 24 and 30 (inclusive), a memory complaint, objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia.

3. Mild AD subjects: MMSE scores between 20 and 26 (inclusive), CDR of 0.5 or 1.0, and meeting NINCDS/ADRDA [46] criteria for probable AD.

3.4 Conclusion

This chapter has provided an overview about the different imaging modalities used in the filed of neuroimaging for the diagnosis of AD. Special attention has been given in this chapter to the structural magnetic resonance imaging (sMRI) because it is widely used in the clinical evaluation of patients with AD. Finally, the last section of this chapter described the tomographic imaging dataset which is used in this PhD thesis.

Chapter 4

Image analysis methods

Recent advances in neuroimaging have made possible to obtain three-dimensional anatomical and metabolic information about the internal structure of the human brain for the diagnosis step. The complexity of brain structures and the differences between brains of different subjects make necessary a process known as preprocessing to prepare the image dataset. Afterwards, a feature selection and feature extraction approaches will be performed to extract the most relevant information from these images before the application of classification process.

4.1 Image pre-processing

4.1.1 Registration or spatial normalization

The raw sMRI images as collected are not yet ready for analysis. Since the analysis methods so far are mostly based on the time series of voxels, the consistency of brain volume position over time is very important. For instance, if the head of the subject moves during data collection, the time series of a voxel may actually consists of several segments from different "real" voxels. Unfortunately, despite efforts on the head fixation, there will always be head motion in sMRI experiments. The preprocessing step to eliminate such motion effects is called *"motion correction"*. Another problem is that the shape and the size of individual brains are different from person to person. For inter-subject studies, we need to have a map from each individual brain space to a standard reference space (template). This process of spatially transforming data into a common space for analysis is known as *inter-subject registration* or *spatial normalization*. The *"Motion correction"* and the *"spatial normalization"* processes use similar technologies.

There exist a number of pieces of software widely used for registering images, such as FreeSurfer [65] or FSL (in the FLIRT and FNIRT package) [66]. They perform linear, non-rigid and elastic transformations or a combination of these. In the studies of this thesis, all the MRI images were registered and spatially normalized using the Statistical Parametric Mapping (SPM) 8 Software [18] yielding a $121 \times 145 \times 121$ voxel representation for each subject. This method; called "cost function", assumes a general affine transformation model with 12 parameters (3 translations, 3 rotations, 3 shears and 3 zooms, as shown in figure 4.1) and a Bayesian framework that maximizes the product of the prior function (which is based on the probability of obtaining a particular set of zooms and shears) and the likelihood function, derived from the residual squared difference between the template and the source image [18]:

$$CF = \sum_i (f(\mathbf{M}\mathbf{x}_i) - g(\mathbf{x}_i))^2 \tag{4.1}$$

where f denotes the source image and g is the template. For each voxel $\mathbf{x}=(x_1, x_2, x_3)$ in an image, the affine transformation into the coordinates $\mathbf{y}=(y_1, y_2, y_3)$ is expressed by a matrix multiplication $\mathbf{y} = \mathbf{M}\mathbf{x}$ [67, 68].

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix}. \tag{4.2}$$

This matrix multiplication is performed globally, as it transforms the whole image, not accounting for local geometric differences. We give an example of the parameters that are computed for scale, translation and shear in 3D, in the following equations:

$$Scale = \begin{pmatrix} m_{11} & 0 & 0 & 0 \\ 0 & m_{22} & 0 & 0 \\ 0 & 0 & m_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.3}$$

$$Translation = \begin{pmatrix} 1 & 0 & 0 & \Delta x_1 \\ 0 & 1 & 0 & \Delta x_2 \\ 0 & 0 & 1 & \Delta x_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.4}$$

$$Shear = \begin{pmatrix} 1 & h_{xy} & h_{xz} & 0 \\ h_{yx} & 1 & h_{yz} & 0 \\ h_{zx} & h_{zy} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.5}$$

In the studies of this thesis, all the MRI images were spatially normalized using the SPM8 yielding a 121 × 145 × 121 three-dimensional structural map for each subject and these images are co-registered to the Montreal Neurological Institute (MNI) space [69].

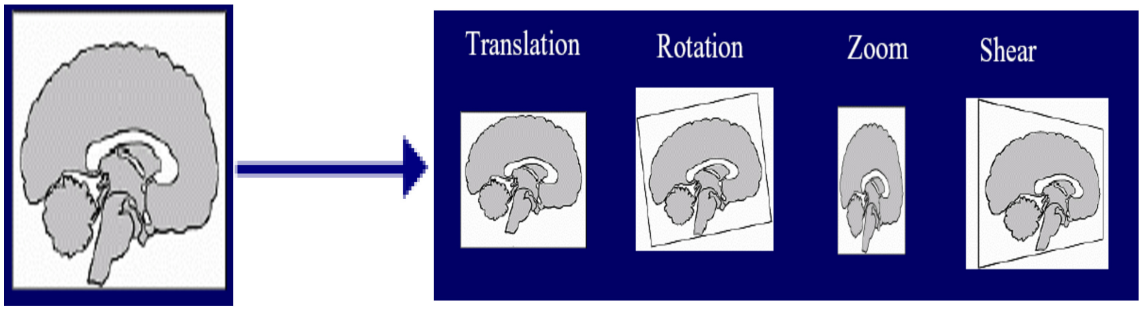


Figure 4.1: Spatial normalization process to match the size and the position of the brain image [70].

This latter is the most widely used coordinate system, recently adopted by the International Consortium for Brain Mapping (ICBM) as its standard template. The

three-dimensional coordinate system defined in MNI was intended to replace the Tailarach space, an older system based on a dissected brain, that was used to compose an atlas by Tailarach and Tournoux [71]. The current template is known as ICBM152, and features the average of 152 normal MRI scans matched to a previous version of the MNI template using a nine parameter affine registration.

4.1.2 Segmentation

Segmentation is an important task in medical image processing and computer aided diagnostic with many applications like detection of morphology [72] or 3-D visualizations for surgical planning [73, 74]. When using MRI images in this thesis, we often refer to the different tissues (grey matter (GM) and white matter (WM) maps), which are the result of the segmentation of the original image. The principal goal of the MRI segmentation process is to partition the brain volume into the mayor tissue classes: gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) [75]. In this thesis, we have used the voxel-based morphometry (VBM) toolbox of the Statistical Parametric Mapping 8 (SPM8) software, which yields GM, WM and CSF maps [76]. It features an Expectation-Maximization (EM) algorithm to model the distribution of the tissue classes as a mixture of gaussians and, by combining this distribution-based information with tissue probability maps using a bayesian rule, the software produces joint posterior probability maps for each tissue. To clean up the segmentation maps, a series of iterative dilations and erosions are used. Finally, since brain regions are expanded or contracted at the spatial normalization step, we can scale the segmented maps using modulation, producing final maps where the total amount of grey matter is preserved.
Figure 4.2 and figure 4.3 show an example of healthy and AD brain GM and WM images respectively. The histogram which is presented in each figure give an idea about the intensity values in each subject according to the number of voxels.

4.2 Feature selection

Feature selection is crucial as a data preprocessing step. In particular when the number of features highly exceeds the number of available samples. This situation is commonly referred to the curse of dimensionality [77], which leads to the decrease of the accuracy rates for the considered learning algorithm. This curse of dimensionality is a major obstacle in machine learning and data mining. Thus, the goal of the feature selection step is to solve the sample size problem by selecting the smallest subset of features that maximally increases the performance of the classification system [78–80]. Therefore, irrelevant features are discarded, and resultant models are faster and more cost-effective.

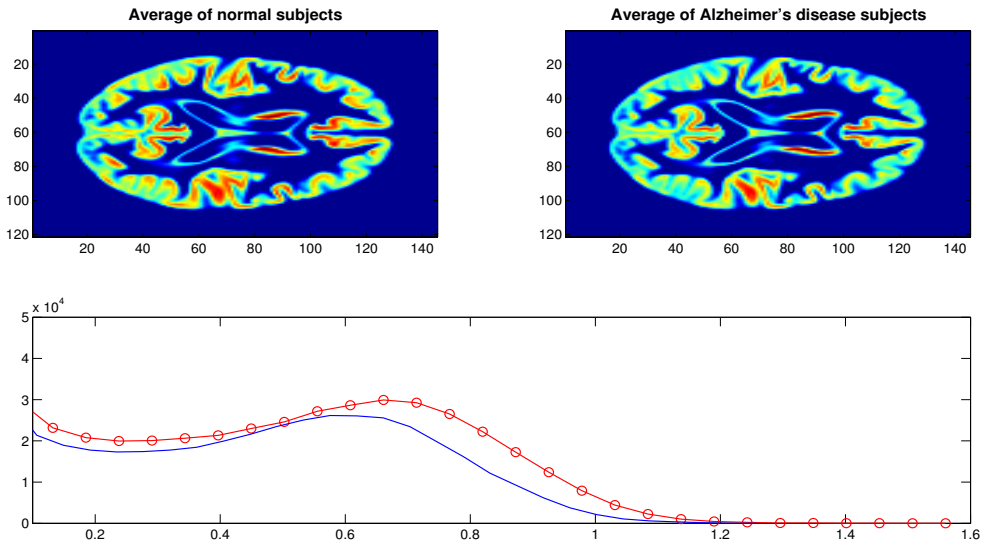


Figure 4.2: The first row: The average of healthy and AD brain images from GM segmented sMRI images. The second row: histogram of the intensity values of each subject according to the number of voxels.

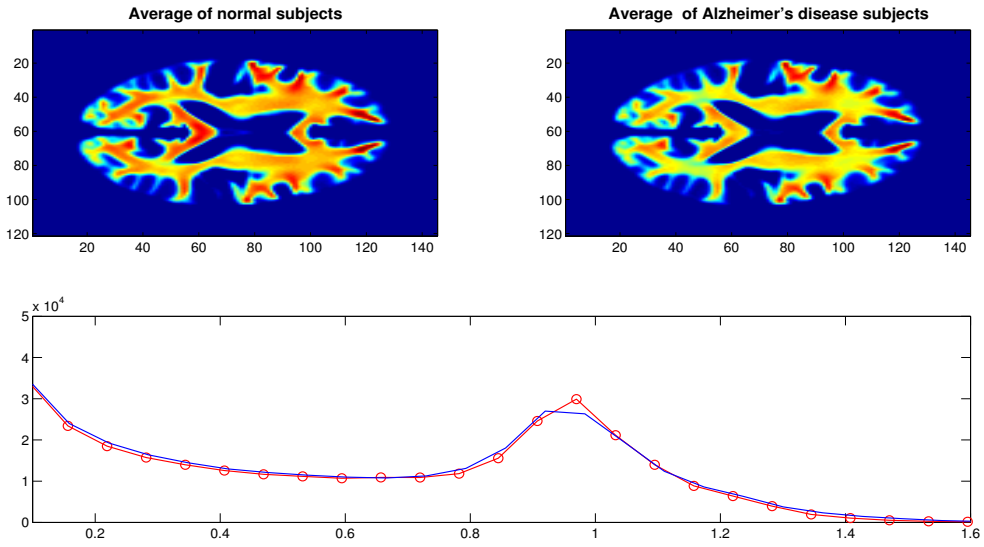


Figure 4.3: The first row: The average of healthy and AD brain images from WM segmented sMRI images. The second row: histogram of the intensity values of each subject according to the number of voxels.

Feature selection can be used before or after feature extraction. When using computationally intensive algorithms such as PCA or ICA, the selection of best features

prior to the decomposition is key to obtain high performance while keeping the computation times small [81, 82]. This also removes noise in some cases where the decomposition algorithm cannot correctly compute the variance. Several techniques of feature selection exist in the literature, such as, t–test, Mann-Whitney-Wilcoxon and relative entropy. However, in this work, only t–test technique is used.

4.2.1 t–test technique

The t-test is a statistical test technique, widely used as a feature selection method [79, 83, 84]. In this thesis, the independent two-sample t-test [85] is used. It quantifies the difference between two classes assuming independent variances. This technique returns a t-value which computes the mean differences between the two classes $\bar{X}_1^f$ and $\bar{X}_2^f$. It uses a common estimation of variance for both classes and assumes normal variables. Let X_i^f a vector containing the f-th feature of all elements in class i ([19, 86, 87]). The t-score of the f-th feature can be computed as:

$$t_f = \frac{\bar{X}_1^f - \bar{X}_2^f}{\sqrt{\frac{\sigma^2_{X_2^f} + \sigma^2_{X_1^f}}{n}}} \tag{4.6}$$

where $\sigma^2_{x_i^f}$ is the variance and $\bar{X}_i^f$ is the average of the f-the feature within class i. The t-test is extensively used in the neuroimaging community, and it is the basis for the SPM and VBM analyses [18].
Figure 4.4 and figure 4.5 show the brain image t_f of GM and WM levels, respectively, with the statistical value of the t-test in each voxel. In these examples, normal and AD images were considered in the calculation of the image t_f. This technique gives us information about voxel class separability. It selects the voxels that present a t-statistic greater than a given threshold ϵ. After the t-test process, the voxels of the brain 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). Thus, the t-test selects the most discriminative features between clinical groups. Then, those selected features will be modeled using feature extraction strategies.

4.2.2 Mann-Whitney-Wilcoxon

Mann-Whitney-Wilcoxon (MWW) rank test, also known as U-test, assigns a rank to all values in the vector corresponding to the f-th feature, X^f, without considering any class. The method used to assign a rank is the 'average', which means that each

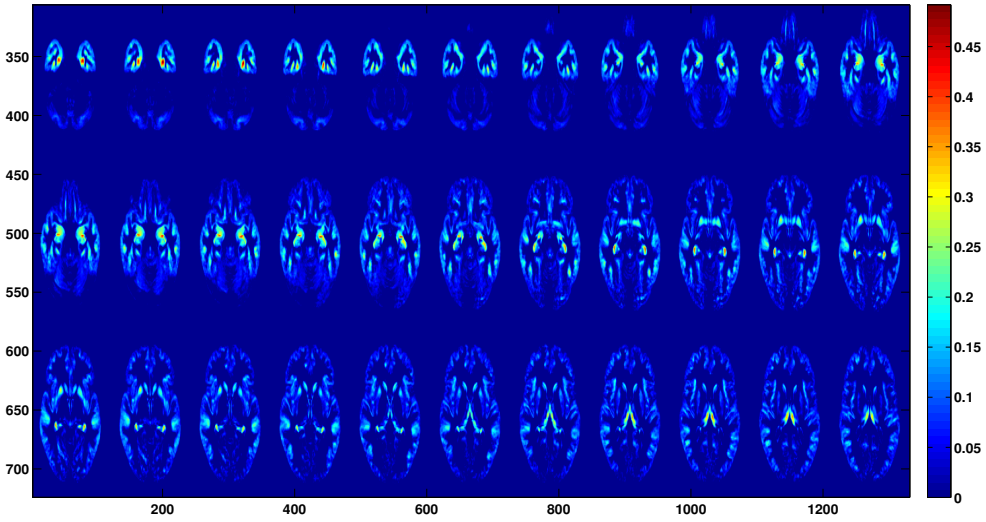


Figure 4.4: Averaged differences in the probabilities of belonging to GM between AD and Normal groups. Color bar represents these differences (maximal difference= 0.5).

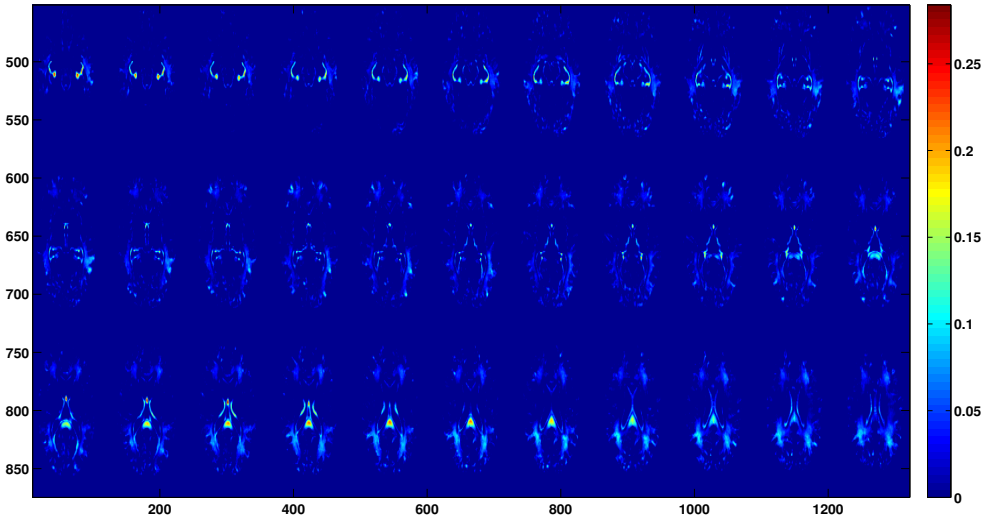


Figure 4.5: Averaged differences in the probabilities of belonging to WM between AD and Normal groups. Color bar represents these differences (maximal difference= 0.3).

value is assigned with the average of the ranks that would have been assigned to all the tied values. Calculation of U value is done by the following expressions [85]:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2} \tag{4.7}$$

where n_1 is the number of elements in class 1, and R_1 is the sum of the ranks in class 1. An equally valid formula for "U" using the sample set of class 2 is:

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2} \tag{4.8}$$

The sum of the two values is given by:

$$U_1 + U_2 = R_1 - \frac{n_1(n_1+1)}{2} + R_2 - \frac{n_2(n_2+1)}{2} \tag{4.9}$$

By taking into account that $R_1 + R_2 = N(N+1)/2$ and $N = n1 + n2$, we find that the sum is:

$$U_1 + U_2 = n_1 n_2 \tag{4.10}$$

We obtain two different values from Eqs. 4.7 and 4.8: U_1 and U_2. The final value of U is taken as the minimum between U_1 and U_2; $U = min(U_1, U_2)$.
This statistical test measures the dissimilarity between two groups of values, and, although is similar to t-test, is less likely than it to spuriously indicate significance because of the presence of outliers.

4.2.3 Relative entropy

The relative Entropy (RE), also known the Kullback-Leibler (KL) divergence is a non-symmetric measure of the difference between two probability distributions Ω_1 and Ω_2. Because of its non-symmetric property, we can make use of this to evaluate the difference between two classes images for each voxel. RE can be calculated with the following equation 4.11 [88]:

$$KL_{\omega_1\omega_2} = \int_V \omega_1 \log \frac{\omega_1}{\omega_2} d\mu \tag{4.11}$$

where μ is any measure of V, the set of all voxels that are placed on a certain brain coordinate, in which $\omega_1 = \frac{\Omega_1}{d\mu}$ and $\omega_2 = \frac{\Omega_2}{d\mu}$ exist. Ω_1 and Ω_2 are two discrete random variables.

4.3 Feature extraction

The feature selection algorithms presented above will perform a significant feature reduction, from hundreds of thousands of voxels to a few thousands. These few thousands voxels are considered the best in discriminating between healthy and affected subjects. The feature selection strategy can be thought of as a mask, in which only the most relevant regions according to the tests are selected. However, this number of features is still large, and therefore, further feature extraction can be applied to extract the most features in the masked regions. In this sense, we have used four algorithms in our CAD systems: principal component analysis, independent component analysis, partial least square and non negative matrix factorization.

4.3.1 Principal Component Analysis

4.3.1.1 Basic concept

Principal component analysis (PCA) is a simple approach to extract the most relevant information contained in image datasets and to use these information for encoding and comparing individual images. Therefore, PCA is applied to neurological images in order to find the principal components (PCs) or, in other term, the eigenvectors of the covariance matrix of the dataset by treating each image as a vector in a high dimensional space. Furthermore, the eigenvectors are ordered according to the amount of variation between the representational images.

4.3.1.2 Mathematical details

The dataset consists of m brain images (of 3 Dimensions), whose typical size is $n = 121 \times 145 \times 121$ voxels. It is understood in this context as a set of column vectors $\mathbf{\Gamma}_i \in \mathbf{R}^n$, $i = 1, 2, ..., m$, formed by concatenating the image voxels.
Thus, $\mathbf{\Gamma}_i^T = (\Gamma_1, \Gamma_2, ..., \Gamma_m)_i$, where Γ_j represents the value of the corresponding voxel intensity j, $j = 1, 2, ..., m$.

The Karhunen-Loeve transformation:

The Karhunen-Loeve Transform (KLT) (also known as Hotelling Transform and Eigenvector Transform) is related to PCA and it is widely used for the analysis of data in several fields, especially, for compressing data in image processing.
Let $\mathbf{\Gamma} \in \mathbf{R}^n$ is n–dimensional vector, there is an accurate representation of this latter ($\mathbf{\Gamma}$) through a set of n linearly independent vectors $\mathbf{u}_i \in \mathbf{R}^n$ as:

$$\mathbf{\Gamma} = \sum_{i=1}^{n} z_i \mathbf{u_i} \tag{4.12}$$

where it is assumed that $\mathbf{u_i}$ vectors are subject to the condition of orthogonality:

$$\mathbf{u_i^T u_j} = \delta_{ij} \tag{4.13}$$

where δ_{ij} is the Kronecker delta. Thus, the equation 4.12 does not describe another thing but only a change of the coordinates to a new orthonormal basis of $\mathbf{R}^n$ where the coordinates of the vector $\mathbf{\Gamma}$ in the new base are given by:

$$z_i = \mathbf{u_i^T \Gamma} \tag{4.14}$$

This coordinate z_i call the i-th component in the new space generated by the base $\mathbf{u}_i$. Let's suppose that in case of a faithful representation of $\mathbf{\Gamma}$ as in eq. (4.12), we are interested to approximate $\mathbf{\Gamma}$ using a small number ($m < n$) of basis vectors $\{\mathbf{u}_i\}$. In this sense, some components z_i whose values are not calculated, will be replaced by random constants b_i. So, the next approximation of $\mathbf{\Gamma}$ is constructed as follow:

$$\hat{\mathbf{\Gamma}} = \sum_{i=1}^{m} z_i \mathbf{u}_i + \sum_{i=m+1}^{n} b_i \mathbf{u}_i \tag{4.15}$$

The approximation error of $\mathbf{\Gamma}$ is giving by the following equation:

$$\begin{aligned} \Delta\mathbf{\Gamma} =& \mathbf{\Gamma} - \hat{\mathbf{\Gamma}} \\ =& \sum_{i=1}^{n} z_i \mathbf{u}_i - \sum_{i=1}^{m} z_i \mathbf{u}_i - \sum_{i=m+1}^{n} b_i \mathbf{u}_i = \\ =& \sum_{i=m+1}^{n} (z_i - b_i)\mathbf{u}_i \end{aligned} \tag{4.16}$$

In order to enhance the approximation error of this approach, a least squares criterion to obtain an optimal solution is established by finding the value of the constant b_i that minimizes the mean square error (MSE):

$$MSE = E\left\{\Delta\mathbf{\Gamma}^2\right\} = \sum_{i=m+1}^{n} E\left\{(z_i - b_i)^2\right\} \tag{4.17}$$

Therefore, the minimization of MSE is equivalent to finding a solution to:

$$\frac{\partial}{\partial b_i} E\left\{(z_i - b_i)^2\right\} = -2(E\{z_i\} - b_i) = 0 \tag{4.18}$$

which simply leads to:

$$b_i = E\{z_i\} = \mathbf{e}_i^T E\{\mathbf{\Gamma}\} \tag{4.19}$$

By the determination of the constant b_i (eq. 4.19), it can be obtained the expected value of z_i components. After that, the MSE can be rewriting as follow:

$$\begin{aligned} MSE &= \sum_{i=m+1}^{n} E\left\{(z_i - E\{z_i\})^2\right\} = \\ &= \sum_{i=m+1}^{n} \mathbf{e}_i^T E\left\{(\mathbf{\Gamma} - E\{\mathbf{\Gamma}\})\right\} E\left\{(\mathbf{\Gamma} - E\{\mathbf{\Gamma}\})\right\}^T \mathbf{e}_i = \\ &= \sum_{i=m+1}^{n} \mathbf{e}_i^T \Sigma_{\mathbf{\Gamma}} \mathbf{e}_i \end{aligned} \tag{4.20}$$

where $\Sigma_{\mathbf{\Gamma}}$ is the covariance matrix of $\mathbf{\Gamma}$. It can be shown in [89, 90] that the optimal choice for $\mathbf{u}_i$ when the following equation is satisfied:

$$\Sigma_{\mathbf{\Gamma}} \mathbf{u}_i = \lambda_i \mathbf{u}_i \tag{4.21}$$

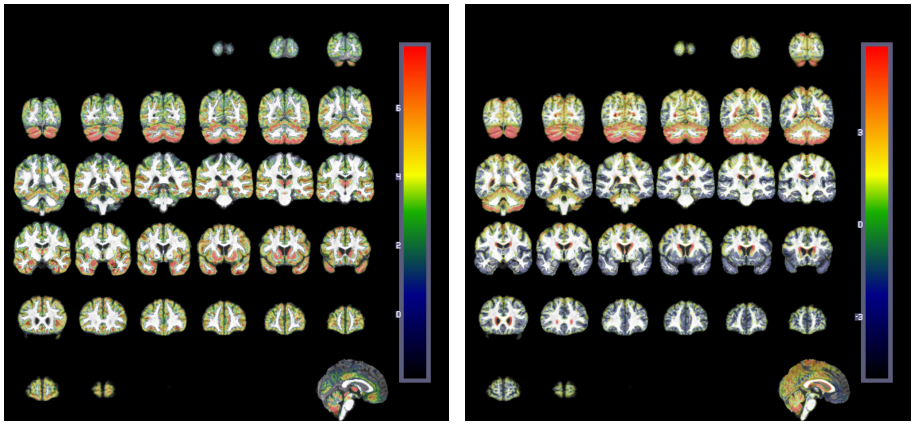
or in other words, when $\mathbf{u}_i$ and λ_i are the eigenvectors and the eigenvalues of the covariance matrix, respectively.

Dimensionality reduction by selection of CPs: Eigenbrains

In brain image datasets, the eigenvectors of the covariance matrix can be understood as a set of vectors that characterize the variation between the images. The location of each brain image contributes more or less to each eigenvector. Thus, this latter can be represented each eigenvector as a self-image (the origin image). The autovectors of the covariance matrix of a set of images can be understood as a set of vectors that characterize the variation between the images. The location of each image contributes more or less to each autovector, so that each autovector can be represented as an own image.

In the field of face recognition [91], the eigenfaces term are used to denote the eigenvalues of the covariance matrix, call "eigenbrains" for the eigenvectors $\mathbf{u}_i, i = 1, ..., N$ [22]. Each brain image can be represented exactly in term of a linear combination

(a) First eigenbrain (b) Second eigenbrain



(c) Third eigenbrain (d) Fourth eigenbrain

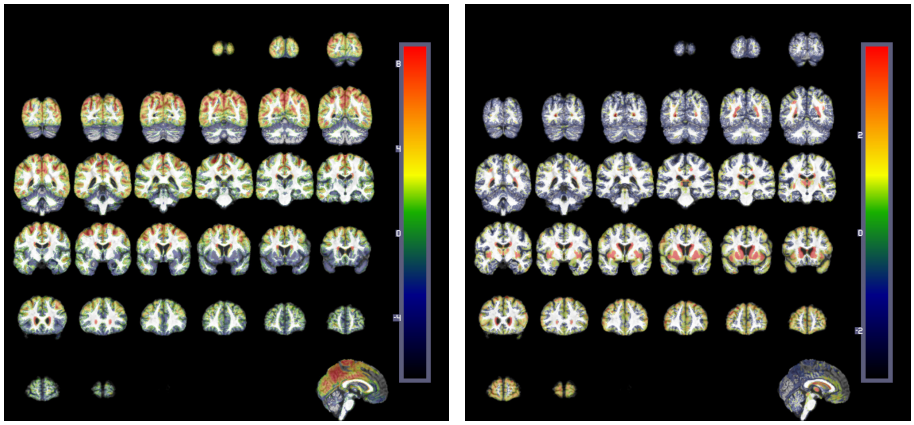


Figure 4.6: The first four Eigenbrains extracted from the ADNI database. They represent the principal components where original images will be projected onto to obtain a dimension reduction.

of the eigenbrains. Furthermore, it can be also approximated using only the "best" eigenbrains which are considered as those that explains most of the variance in the brain image dataset.
Figure 4.6 shows the first four eigenbrains obtained with a segmented sMRI reference image (gray level) by the PCA technique. The best m eigenbrains extend a subspace with m– dimensions of all the possible images. Therefore, each individual can be characterized by the small set of weights associated with the eigenfaces needed to describe it and to rebuild it. This is an extremely compact representation

when compared with the same eigenfaces.

- ***Effective computation of the Eigenbrains:***

After the preprocessing step, we get a 3D representation of each subject with size $n = 121 \times 145 \times 121$ voxels. $\mathbf{I} = [\mathbf{I_1}, \mathbf{I_2}, ..., \mathbf{I_N}]$ is the representation of all the dataset that is rearranged in form of n vectors, where N is the number of samples.
The average of these vectors is defined as:

$$\mathbf{I}_M = \frac{1}{N} \sum_{i=1}^{N} \mathbf{I}_i \tag{4.22}$$

After the subtraction of the averaged vectors from the segmented brain image vectors, a new set of vectors is obtained; $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}]$. Where each $\mathbf{x_i}$ represents a vector of n-dimensions, $\mathbf{x_i} = (x_{i1}, x_{i2}, ..., x_{in})^T$, $i = 1, 2, ..., N$. The covariance matrix is defined as follows:

$$\sum \mathbf{X} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i {\mathbf{x}_i}^T = \frac{1}{N} \mathbf{X}\mathbf{X}^T \tag{4.23}$$

$\sum \mathbf{X}$ is a matrix of $n \times n$ dimensions, and the determination of these n eigenvectors and these n eigenvalues can become a difficult task due to the typical size n. However, there is a reliable method to find these eigenvectors. Note that if the sample size N is smaller than the input space of n dimension, an exist only N-1 (in place of n) significant eigenvectors. The resting eigenvectors will be associated with eigenvalues of zero value. Fortunately, the problem of eigenvectors can be solved by first solving the eigenvectors of $N \times N$ matrix and then by taking appropriate linear combinations of $\mathbf{x}_i$ images. Consider that $\mathbf{v}_i$ are the eigenvectors of the matrix $\mathbf{X}\,\mathbf{X}^T$, that is defined as follow:

$$\mathbf{X}^T \mathbf{X} \mathbf{v}_i = \mu_i \mathbf{v}_i \tag{4.24}$$

Multiplying both sides of the equation 4.24 by the matrix $\mathbf{X}$, the following equation is obtained

$$\mathbf{X}\mathbf{X}^T \mathbf{X} \mathbf{v}_i = \mu_i \mathbf{X} \mathbf{v}_i \tag{4.25}$$

where $\mathbf{X}\mathbf{v}_i$ are the eigenvectors of $\sum \mathbf{X} = \mathbf{X}\mathbf{X}^T$. The matrix $\mathbf{L} = \mathbf{X}^T\mathbf{X}$ of dimension $N \times N$, where $\mathbf{L}_i^j = \mathbf{X}_i^T \mathbf{X}_j$, and the eigenvectors, $\mathbf{v}_l$ are calculated from $\mathbf{L}$. These

vectors are determined by a linear combination of the set of N training images to form $\mathbf{u}_l$ eigenbrains.

$$\mathbf{u}_l = \sum_{k=1}^{N} \mathbf{v}_{lk}\mathbf{x}_k \tag{4.26}$$

With the above analysis, the calculation are reduced in function of the number of voxels n and in function of the number of the training set images N. The size of the training set will be relatively small ($N \ll n$). Thus, the calculation will be more manageable. In addition, the eigenvalues allow us to order their associated eigenvectors according to their utility to characterize the variation between images.

- ***The use of the eigenbrains for the classification step:***

In practice, only a number $M \ll N$ is sufficient for eigenbrains image classification, since no accurate reconstruction is required. In this context, the classification becomes a pattern recognition task. The m eigenbrains make up a subspace of m-dimensions of the original input space of n-dimensions. The most significant eigenvectors m of the $\mathbf{L}$ matrix are chosen in principle as those with higher associated eigenvalues. A new image $\mathbf{I}$ is transformed into eigenbrains components or coefficients PCA to project it into the "brain" space by a simple operation

$$w_k = \mathbf{u}_k^T(\mathbf{I} - \mathbf{I}_M) \tag{4.27}$$

where $k = 1,...,m$, the Weight w_k form a vector $\Omega^T = [w_1, w_2, ..., w_m]$ that describe the contribution of each eigenbrain in the representation of the input image. Therefore, the weight vector can be used within a set of predefined classes the best description of the image, in a standard pattern recognition algorithm.

4.3.2 Independent Component Analysis

4.3.2.1 Basic concept

Independent Component Analysis (ICA) [92], is a statistical technique that represents a multidimensional random vector as a lineal combination of non-gaussian random variables (the so-called "independent components") to be as independent as possible. ICA has been used widely on segmentation and clustering of medical images [17, 93]. It can be considered as a non-gaussian version of Factor Analysis (FA) [81]. This algorithm is used as a new strategy to avoid the small sample size

problem [23]. This problem occurs when the number of input features to the classifier is higher than the number of samples used to train this classifier.
In this current work, the number of samples is around 818 images, and the number of selected voxels N is around 8000. Thus, a reduction in the input vector is desirable. Using ICA, we obtain a representation of the selected voxels in the IC space, where the number of components K is equal to 8. This way, we ensure that the system is not affected by the small sample size problem.

4.3.2.2 Mathematical details

Assume that we observe n lineal mixtures $x_1, x_2, \ldots, x_n$ of length N that can be modeled as an expression of K independent components (IC). These independent components are defined as $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \ldots \mathbf{s}_K)$, where each $\mathbf{s}_K$ vector has a length of N. So, each random vector x_n can be described as a linear combination of K independent components:

$$\mathbf{x_n} = a_{1n}\mathbf{s}_1 + a_{2n}\mathbf{s}_2 + \ldots + a_{Kn}\mathbf{s}_K \tag{4.28}$$

Without loss of generality, we can assume that both the observed vectors and the independent components are zero mean. If the previous conditions are not met, the x variables can be centered by subtracting the sample mean. To use a vector-matrix notation, more convenient in this case, we denote as matrix X the random vector whose elements are $x_1, \ldots, x_n$. We also denote as A the matrix that contains all a_{Kn} elements, the "mixing matrix" that projects each image into the space defined by the IC. Using this notation, the mixing model above remains as follows:

$$\mathbf{X} = \mathbf{AS} \tag{4.29}$$

Figure 4.7 presents the first four ICs obtained using this proposed method applied to the segmented sMRI database when AD and NC subjects are considered. These components have been spatially represented by assigning the brain coordinate of each voxel to each of the values in these IC space. It can be observed that each component highlights different zones that are usually related to AD. The starting point of ICA is the assumption that all components s_K are statistically independent. To measure independence, we assume that all independent components have a non-gaussian statistical distribution. It is assumed that a sum of independent signal trends to gaussianity, so if non-gaussianity is maximized with any independence criteria F, for instance, the kurtosis or negentropy, we obtain signals that are more independent than the previous ones [92, 94]. After estimating the matrix A, we can compute its inverse, W and obtain the projection $\mathbf{S}$ of the images in the dataset into the IC space with:

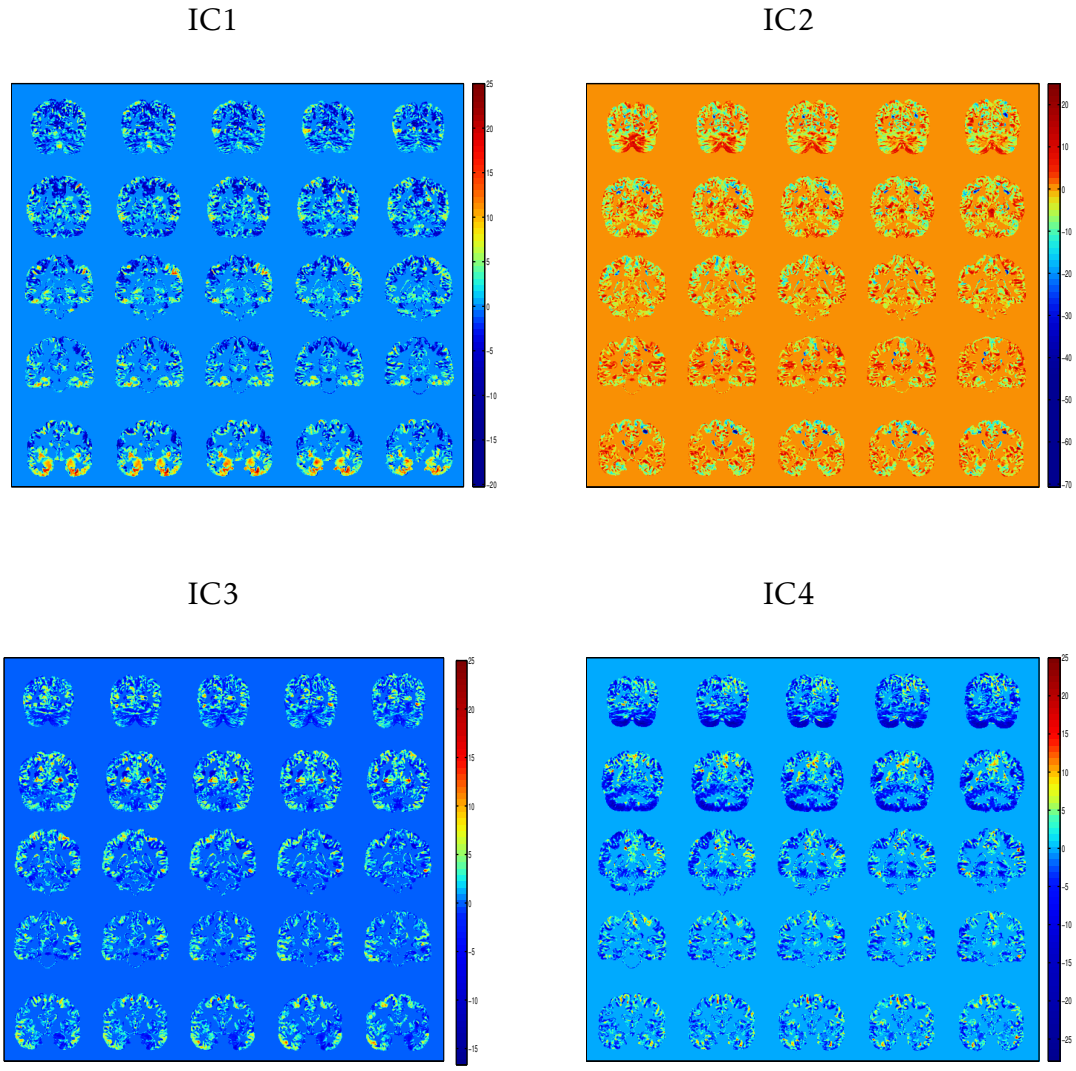


Figure 4.7: Representation of the first four components obtained by ICA in the ADNI database.

$$\mathbf{S} = \mathbf{WX} \tag{4.30}$$

- ***FastICA:***

Adaptive algorithms based on gradient descend can be problematic when they are used on an environment in which adaptation is not necessary, like this case. The convergence is often slow, and depends on the choice of convergence parameters. As a solution to this problem, block algorithms based on fixed-point iteration [95, 96]

can be used. In [95], a fixed-point algorithm based on kurtosis is introduced. In [96], this algorithm, known as FastICA, is generalized to general contrast functions. The single unit FastICA algorithm has the following form:

$$\mathbf{w}(k) = E\{\mathbf{x}g(\mathbf{w}(k-1)^T\mathbf{x})\} - E\{g'(\mathbf{w}(k-1)^T\mathbf{x})\}\mathbf{w}(k-1) \tag{4.31}$$

where the loadings vector $\mathbf{w}$ is normalized to unit norm in each iteration, and the function $g(x)$ is a derivative of the contrast function G defined in [94]. The expected values are estimated in practice by using the mean of a significantly high number of samples of the input data. The speed of convergence of the fixed-point algorithms is clearly superior to more neural algorithms. Improvements between 10 and 100 times in the speed are observed frequently [97]. FastICA has been used in this thesis work to perform the ICA approach.

4.3.3 Partial Least Squares

4.3.3.1 Basic concept

The partial least squares (PLS) [98, 99] approach was developed in 1975 by Herman Wold statesman [100] for the treatment of material chains trices and economic applications. Later, his son and other researchers extended the PLS idea to other fields. In the recent years, the interest of PLS technique has increased specially in Regression models such as Principal Component Regression (PCR) [101]. The effectiveness of PLS has been studied theoretically in terms of its variance [102] and its properties of contraction [103]. In the same way, the performance of PLS has been studied in several papers [104–106].
In its general form, PLS creates orthogonal score vectors by maximizing the covariance between the different sets of variables. This score vectors called also the latent vectors or the components [107]. Furthermore, PLS can be naturally extended to regression problems.

- ***The principal steps of the regression system are following:***

First, the observed variables, X, are converted into an intermediate set of latent variables and these new variables are used for the regression step with the taking into account the criteria set Y. Furthermore, the most criteria used for the calculation of the latent vectors is the covariance between the maximum scores in Y (or between scores in X and scores in Y). This approach combines the high variance of X and the high correlation [98].

- ***The main concepts of PLS are following:***

– PLS can easily be extended as a linear regression method very powerful and it can work with a large number of variables.

– The resulting model predicts one or more properties, Y, from the original dependent variables, X.

– During the development, a small calculated number of PLS components are used internally for the regression.

– The number of PLS components determine the complexity of the model and it can be optimized to increase the predictability of the algorithm.

PLS can be also applied to solve the classification problems by using a coding matrix of a vector belonging for a determinate class. A classifier based on PLS is, in some ways, similar to a linear classifier [108]. Moreover, the powerful machinery of kernels based learning can be also applied to PLS approach [109]. A general model of PLS approach is applied in this work as a discriminatory tool and dimension reduction method by using the latent vectors as feature vectors. After the extraction of the feature vectors, an appropriate classifier will be applied.

4.3.3.2 Mathematical details

PLS is often described as a numerical algorithm that maximizes an objective function under certain conditions. This objective function is the covariance between the score vectors **X** and **Y**, while the condition is usually the orthogonality of these vectors. Furthermore, the margins of the objective function which is used, distinguishes between two variants of PLS, known as PLS1 (when **Y** is a vector) and PLS2 (when both **X** and **Y** are multidimensional variables).
The PLS2 is a generalization of PLS1. It is assumed that both **X** and **Y** are matrices of size $(n \times m)$ and $(n \times q)$ respectively. Each row of **X** corresponds to an observation which is represented by characteristics m (predictor variables). For its part, the matrix **Y** collected q properties of each one of the n observations. To facilitate the notation, we will assume that the columns of **X** and **Y** are centered in the middle.
The goal of PLS2 regression is to find a linear relationship:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{4.32}$$

between the variables in **X** and **Y** using a coefficient matrix, **B**, of size $m \times q$ and a error matrix, **E**.

In the case of PLS1, the problem is reduced to

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \tag{4.33}$$

with coefficients **b** and error **e**.
In the case of searching the relationship directly, both, **X** and **Y** are modeled by latent variables according to the following models Regression:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_x \tag{4.34}$$

and

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_y \tag{4.35}$$

With $\mathbf{E}_x$ and $\mathbf{E}_y$ are the error matrices. **T** and **U** (scores matrices) as well as the matrices **P** and **Q** (loadings matrices) have columns, with a (inferior ou equal) min (m, q, n) being the number of PLS components. The values of **T** are a linear combination of the values of **X**, so it can be considered a good summary of **X**. In the same way, **U** is a good summary of **Y**. Furthermore, t_j, u_j, p_j and q_j (j = 1, ..., n) will be used to denote the jth column of **T**, **U**, **P** and **Q** respectively. In addition, the **X**-scores and the **Y**-scores are referred to the values of **T** and **U** matrices respectively. Since, these matrices represent the scores vectors **X** and **Y**.
The values of **T** and **U** are connected by the following linear relationship:

$$\mathbf{U}_j = \mathbf{d}_j \mathbf{T}_j + \mathbf{h}_j \tag{4.36}$$

where $\mathbf{h}_j$ is the residue and $\mathbf{d}_j$ is the regression parameter.
If, for example, the linear relationship between $\mathbf{T}_1$ and $\mathbf{U}_1$ is strong (the elements of h_1 are small). Then, the **X**-scores of the first PLS components predict fairly well the **Y**-scores and therefore also adequately predict the values of **Y**. However, in the most cases, it used more than one component for modeling **Y** from **X**. The relationship between the score matrices is therefore as follows:

$$\mathbf{U} = \mathbf{TD} + \mathbf{H} \tag{4.37}$$

where **D** is a diagonal matrix with d_1, d_2, ..., d_a elements and **H** is the matrix of residues (with $\mathbf{h}_j$ which refers to the columns of **H**).
PLS is motivated by the equation 4.37 as only part of the information **X** and **Y** is used for the regression.
In the case of PLS1, the **Y**-scores are not available and the equation 4.37 is reduced to

$$\mathbf{y} = \mathbf{Td} + \mathbf{h} \tag{4.38}$$

The objective of PLS2 is to maximize the covariance between the scores vectors of **X** and **Y**, while for PLS1, the objective is to maximize the covariance between the **X**-scores and **Y**. The covariance is used as a criterion for calculating the latent variables by combining the high variance of **X** and also the high correlation between **X** and **Y** although this depends largely as estimated covariance. In the classic case, the covariance between two vectors **t** and **u** is calculated as the sample covariance, $\mathbf{t}^{\mathbf{T}}\mathbf{u}/(n-1)$, but also it can be used more robust estimators. Since, the maximization problem can not be unique; a constant need on the score vectors and this is usually $\|\mathbf{t}\| = \|\mathbf{u}\| = 1$ (length 1). The score vectors are obtained from the projection of **X** and **Y** matrices of loadings vectors. Although, it seems logical to use here loading vectors forming the **P** and **Q** matrices (equations 4.34 and 4.34). For technical reasons, we use other loading vectors, specifically, **w** for **X** and **c** for **Y**:

$$\mathbf{t} = \mathbf{Xw} \tag{4.39}$$

$$\mathbf{u} = \mathbf{Yc} \tag{4.40}$$

The constrained maximization problem is as follows:

$$cov(\mathbf{Xw}, \mathbf{Yc}) \rightarrow \begin{cases} \mathbf{t} = \mathbf{Xw} = 1 \\ \mathbf{u} = \mathbf{Yc} = 1 \end{cases} \tag{4.41}$$

Note that the restriction may be that the length of the score vectors either 1 (as has been chosen here) or the weight vectors, **w** and **c** are 1. The solution of the maximization problem are the first score vectors, i.e, $\mathbf{t}_1$ and $\mathbf{u}_1$. For the following score vectors, they have the same criteria, but with additional restrictions. Normally, these new restrictions are the orthogonality of the previous score vectors, ie, $\mathbf{t}_j^T \mathbf{t}_l = 0$ and $\mathbf{u}_j^T \mathbf{u}_l = 0$ for $1 \leqslant j \leqslant l < a$. An alternative strategy is to require the orthogonality of the loading vectors which leads to non-orthogonal score vectors and therefore correlated.
To find the first component and given that we use the sample covariance, the maximization problem can be rewritten as maximizing

$$\mathbf{t}^T\mathbf{u} = \mathbf{Xw}^T\mathbf{Yc} = \mathbf{w}^T\mathbf{X}^T\mathbf{Yc} \rightarrow max \tag{4.42}$$

Under the same length restriction vectors equal to 1. The solutions for **w** and **c** are found easily by means of the decomposition in Securities Singular (SVD) [110] of **XTY**. Accordingly, from all the possible directions of **w** and **c**, the maximum for the equation 4.42 is obtained for the vectors **w**=**w1** and **c**=**c1** corresponding to largest singular value of **XTY** [111].

PLS algorithm:

Different algorithms have been proposed to solve the maximization problem posed in equation 4.42. This section will describe one of the most used algorithms. It is based on the strategy of requiring the orthogonality of vector scores, resulting in score uncorrelated vectors.

- ***Description:***

Proposed by Jung in 1993, this algorithm directly maximize the initial problem (equation 4.41) under the constraint of orthogonality of the scores-t [112]. It was actually derived to solve specific objective function, i.e. to maximize the covariance. In this case, the covariance matrix is $\mathbf{S} = \mathbf{X^T Y}$.
A description of the PLS algorithm is shown in the following pseudo-code:

1- Initialize $\mathbf{S_0} = \mathbf{X^T Y}$ and repeat the steps from 2 to 8 for j = 1, ..., a

2- If j = 1, $\mathbf{S_j} = \mathbf{S_0}$, else $\mathbf{S_j} = \mathbf{S_{j-1}} - \mathbf{P_{j-1}}(\mathbf{P_{j-1}}^T\mathbf{P_{j-1}})^{-1}\mathbf{P_{j-1}}^T\mathbf{S_{j-1}}$

3- Calculate $\mathbf{w_j}$ as the first (left) singular vector of $\mathbf{S_j}$.

4- $\mathbf{w_j} = \frac{\mathbf{w_j}}{\|\mathbf{w_j}\|}$

5- $\mathbf{t_j} = \mathbf{X w_j}$

6- $\mathbf{t_j} = \frac{\mathbf{t_j}}{\|\mathbf{t_j}\|}$

7- $\mathbf{p_j} = \mathbf{X_j}^T\mathbf{t_j}$

8- $\mathbf{p_j} = [\mathbf{p_1}, \mathbf{p_2}, ..., \mathbf{p_{j-1}}]$

The weights $\mathbf{w_j}$ and the resulting scores $\mathbf{t_j}$ are stored as the columns of the matrix $\mathbf{W}$ and $\mathbf{T}$ respectively. Note that the matrix $\mathbf{W}$ differs from that obtained with other algorithms because the weights are related directly with $\mathbf{X}$ and not with the reduced matrices. Step 2 makes the constraint of orthogonality of the $\mathbf{t_j}$ scores compared to previous scores vectors given that the search is performed in the orthogonal complement of $\mathbf{S_{j-1}}$. Step 3 directly maximizes the initial problem (equation 4.41). The scores obtained in step 5 by projecting $\mathbf{X}$ in the optimum direction and the loadings matrices are obtained in step 7 by means of least squares (Equation 4.34). The final coefficients of the equation 4.32 are:

$$\mathbf{B} = \mathbf{W}\mathbf{T}^T\mathbf{Y} \tag{4.43}$$

For PLS1 technique, the algorithm is simplified slightly. In this case, the orthogonality in step 2 is achieved when using the following projection

$$\mathbf{S_j} = \mathbf{S_{j-1}} - \mathbf{P_{j-1}}(\mathbf{P_{j-1}^T}\mathbf{P_{j-1}})^{-1}\mathbf{P_{j-1}^T}\mathbf{S_{j-1}} \tag{4.44}$$

Dimensionality reduction by selection of the relevant PLS-brains:

Once the PLS algorithm is executed, the extraction of feature vectors is trivial. As noted in section 4.3.3.2, PLS algorithm compose of two data sets, **X** and **Y**. Each sets of them is as a product of two matrices which is known as loadings matrix and scores matrix (equations 4.34 and 4.35).
In this work (the case of PLS1), the set **X** is defined as a matrix which contains the tomographic images. While, the set **Y** is a vector that contains the labels of these images. To build **X** from the corresponding three-dimensional matrices of images. First, it is necessary to reduce the dimensionality of each image. Thus, the computation time will be reduced and the effectiveness of the algorithm to eliminate the redundant information will be improved because the marks of Alzheimer tomographic images are not in the voxels level but at the structural level (in the higher level). In order to select only the high intensity voxels of the brain, a binary mask is applied. This mask is constructed by taking the voxels with an intensity higher than 10% of the maximum intensity. After applying the mask, the resulting voxels of each image are put into a vector form remaining the set of all images as a two-dimensional matrix.

The SIMPLS algorithm has been used in this work to carry out the decomposition in scores and loadings matrices (equations 4.34 and 4.35). The input parameters of the algorithm (matrices **X** and **Y**) are constructed as follows: The matrix **X** is formed by the images (a row for each image and a column for each voxel) and the **Y**-matrix is a vector with the labels of **X**-images. Thus, we can perform an interesting analysis based on the concept of eigenbrains through the scores matrices and the loadings matrices. That way, loadings would be viewed as elementary brain images and scores would represent the quantity of loadings used for building a specific image. The PLS-based methodology maximizes the covariance taking into account the labels information, thus the PLS-brains contain the differences between the two classes. In addition, most of the variance is gathered by the first components and therefore, most of the differences are also gathered by the first components.
Figure 4.8 shows the representation of the first four PLS-brain obtained by the PLS algorithm. Furthermore, the feature extraction approach has been performed by the k-fold cross validation strategy in order to avoid the label of an image which is taken into account in the extraction of features for this image. Therefore, for each image of the database, it has implemented the algorithm with the remaining images (**X** contains n-1 and the **Y** labels corresponding to the images in **X**). In each obtained

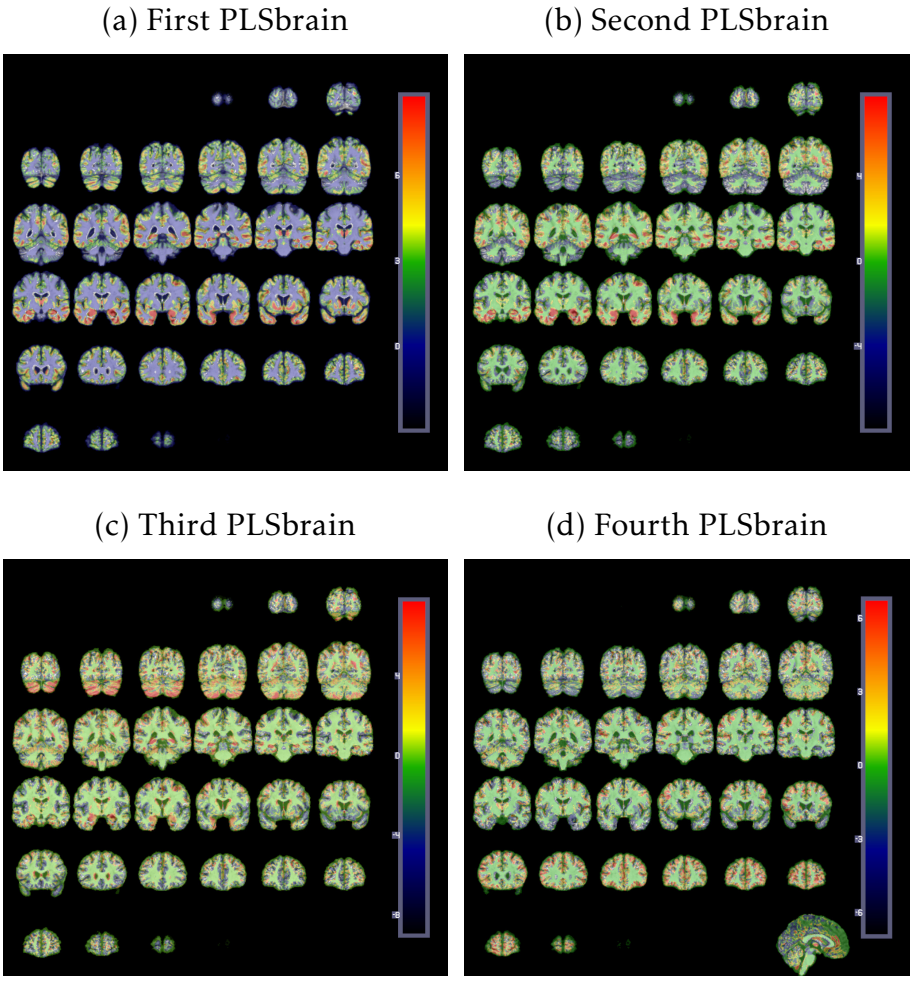


Figure 4.8: Representation of the four PLS-brain of an MRI reference image (gray level) obtained by the PLS algorithm.

execution, in addition to the scores and loadings matrices, a weight matrix with a score vector of the image under study is calculated. This scores vector is used as feature vector.

4.3.4 Non-negative Matrix Factorization

4.3.4.1 Basic concept

Non-negative Matrix Factorization (NNMF) is a recently developed technique for finding reduced linear representations of non-negative data [113–115], used for finding the reduced decomposition tool for multivariate data. This technique is

especially suitable for nonnegative data sets, where all the variables consist of positive values. It has been widely applied in the field of image processing [116, 117] and in particular, it is used in brain image analysis [118].

4.3.4.2 Mathematical details

Mathematically, NNMF is a linear, non-negative approximate data representation where the original database $\mathbf{A} = [\mathbf{A_1}, \mathbf{A_2}..., \mathbf{A_N}]$ (M by N elements), which consists of N measurements of M non-negative scalar variables, is approximated by a non-negative matrix product, as given:

$$\mathbf{A} = \mathbf{WH} \tag{4.45}$$

where, the matrix $\mathbf{W} = [\mathbf{W_1}, \mathbf{W_2}..., \mathbf{W_K}]$ has dimension $M \times K$, and the matrix $\mathbf{H} = [\mathbf{H_1}, \mathbf{H_2}..., \mathbf{H_N}]$ has dimension $K \times N$. Thus, matrix $\mathbf{A}$ is decomposed as depicted in equation 4.46.

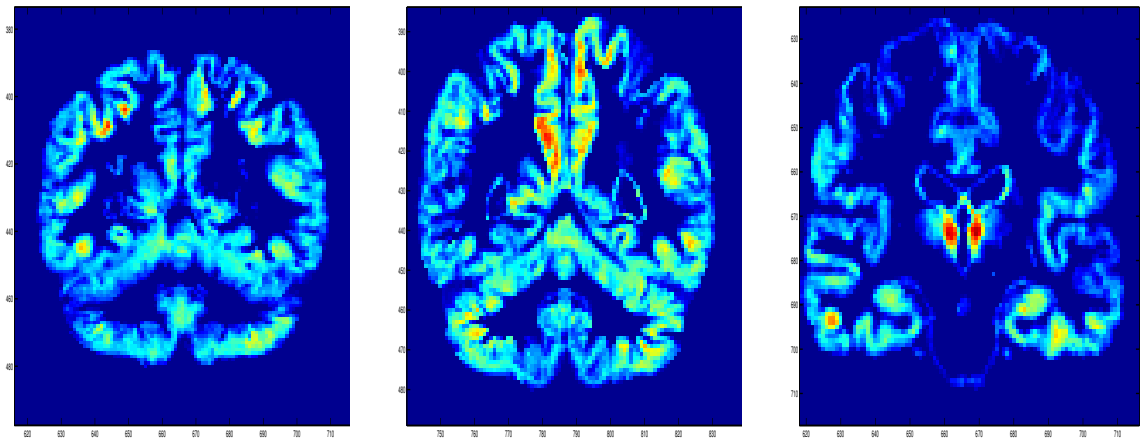
$$\mathbf{A}_{mn} = \sum_{k=1}^{K} \mathbf{W}_{mk}\mathbf{H}_{kn} \tag{4.46}$$

Thus, the NNMF method tries to approximate the data $\mathbf{A}$ by means of a product of matrices $\mathbf{WH}$, where $\mathbf{W}$ is the new space of representation and $\mathbf{H}$ is the representation of the data in the transformed space, according to the space basis. As a consequence, this transformation approach yields to reduce the matrix $\mathbf{H}$ which represents $\mathbf{A}$ in terms of $\mathbf{W}$. An appropriate decision on the value of K is critical in practice, but the choice of K is very often problem dependent. In most cases, K is chosen such that $K \ll \min(N, M)$ in which case $\mathbf{WH}$ can be thought of as a compressed form of the data in $\mathbf{A}$.

After NNMF factorization, the data contained in $\mathbf{H}$ (K by N elements) can be considered as a transformed database with lower number of features (k), than the original database $\mathbf{A}$. As a consequence, a few variables are representing the data of each measure in the new representation.
The *relative error* of the factorization can be computed by means of the comparison of matrix $\mathbf{A}$ and the approximation $\mathbf{WH}$. The minimum number of vectors K in the NNMF basis is selected so that a predefined level of relative error is not exceeded.

- ***Factorization Rule:***

Given the data matrix, the optimal choice of matrices and are defined to be those nonnegative matrices that minimize the reconstruction error between $\mathbf{A}$ and $\mathbf{WH}$. A



(a) feature k=1 (b) feature k=2 (c) feature k=3

Figure 4.9: A transversal brain MRI feature extraction by means of NNMF for one slice of a patient. Image of the k vectors of the new NNMF basis with k=3.

variety of error functions (Error) have been proposed in [113, 115]. One of the most useful is given below, and applied in this work:

$$Error = \frac{1}{NM}\|\mathbf{A} - \mathbf{WH}\|^2 = \frac{1}{NM}\sum_{nm}(\mathbf{A}_{nm} - \mathbf{WH}_{nm})^2 \tag{4.47}$$

This error function is known as the Frobenius norm (reduction of the Euclidean distance). Thus, the NNMF process is translated into an optimization problem, through the minimization of the Error function.

4.4 Conclusion

Pre-processing and post-processing steps are a very critical steps to obtain meaningful results. In this sense, several important processing steps, such us, anatomical segmentation, feature selection and feature extraction techniques are applied in this work.
t-test feature selection technique was used in this work to select the most relevant regions related to AD. This technique is based on the computation of a feature relevance score directly on the dataset. The relevance score is used to sort the relevant features, discarding those with a lower score. The application of feature selection technique before feature extraction technique is the key to obtain high performance while reducing the computation time. Afterwards, several feature extraction tech-

niques, such as, PLS, PCA, ICA and NNMF techniques, were used in the different CAD systems developed in this work. These feature selection techniques were applied to decrease more the number of features by the decomposition of the selected regions. All mathematic details about these techniques have been presented in this chapter. SPM software [119] is applied to perform the MRI image segmentation, which is described in detail in the appendix A.

Chapter 5

Classification and resampling methods for validation

The final step of any pattern recognition system is to learn a model from the training instances capable of correctly classifying future unseen data. As already mentioned in the previous chapters, the high dimensionality of the feature vectors (even after the feature selection and feature extraction steps) suggests the use of a discriminative model. The support vector machine (SVM) classifier [23] is frequently used for classification in medical imaging including computer-aided diagnosis in magnetic resonance brain imaging [120–123]. In this sense, SVM have been used in this work as a discriminative tool for CAD systems to classify the images under study. A comprehensive description of its basic concepts and its mathematics forms will be given in section 5.1. The performance metrics which are used to evaluate the classifier will be presented in section 5.2. On a different topic, the assessment of any CAD system is very important for obvious reasons. In this PhD work, the cross-validation (CV) technique was used to estimate the performance measures such as the classification accuracy, sensibility or specificity. Section 5.3 will present the different CV methods.

5.1 Support Vector Machine

5.1.1 Basic concepts

The Support Vector Machine algorithm was first developed in 1963 by Vapnik and Lerner [124] and Vapnik and Chervonenkis [125, 126] as an extension of the Generalized Portrait algorithm. This algorithm is firmly grounded in the framework of statistical learning theory–Vapnik Chervonenkis (VC) theory, which improves the generalization ability of learning machines to unseen data [126–128].
In the last few years, Support Vector Machines (SVM) have attracted attention from the pattern recognition community [121, 129] owing to a number of theoretical and computational merits derived from Statistical Learning Theory (SLT) [129]. SVM [130] look for the set of support vectors that allow to build the optimal discriminating surface in the sense of providing the greatest margin between the classes [131–133]. It separates a given set of binary labeled training data with a hyperplane that is maximally distant from the two classes (known as the maximal margin hyperplane), as exemplified in figure 5.1.

5.1.2 Mathematics

The simplest support vector classifier assumes the data to be linearly separable. Each sample of a given dataset is considered to be a p-dimensional vector. Therefore, the objective is to separate a set of data, with binary labels, using a hyperplane (multidimensional plane) that maximizes the distance between the two classes. Therefore, we must build a function f: $\mathbb{R}^p \rightarrow \pm 1$ that is able to assign a binary value (+1 or -1) to a new sample. This function is built using a set of training data $\mathbf{x}$, which contains p-dimensional vectors (samples) $\mathbf{x_i}$ and their corresponding class y_i, so that f will correctly classify new examples ($\mathbf{x}$,y):

$$(\mathbf{x_1}, y_1), (\mathbf{x_2}, y_2), ..., (\mathbf{x_N}, y_N) \in \mathbb{R}^p \times \pm 1 \tag{5.1}$$

The linear support vector classifier defines a separation hyperplane in a multidimensional space using the following function:

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + \omega_0 = 0, \tag{5.2}$$

where $\mathbf{x}$ denotes the feature vector, $\mathbf{w}$ is known as the weight vector and ω_0 as the threshold. The decision hyperplane position is determined by the vector $\mathbf{w}$ and ω_0: the vector $\mathbf{w}$ is orthogonal to the decision plane and ω_0 determines its distance to the

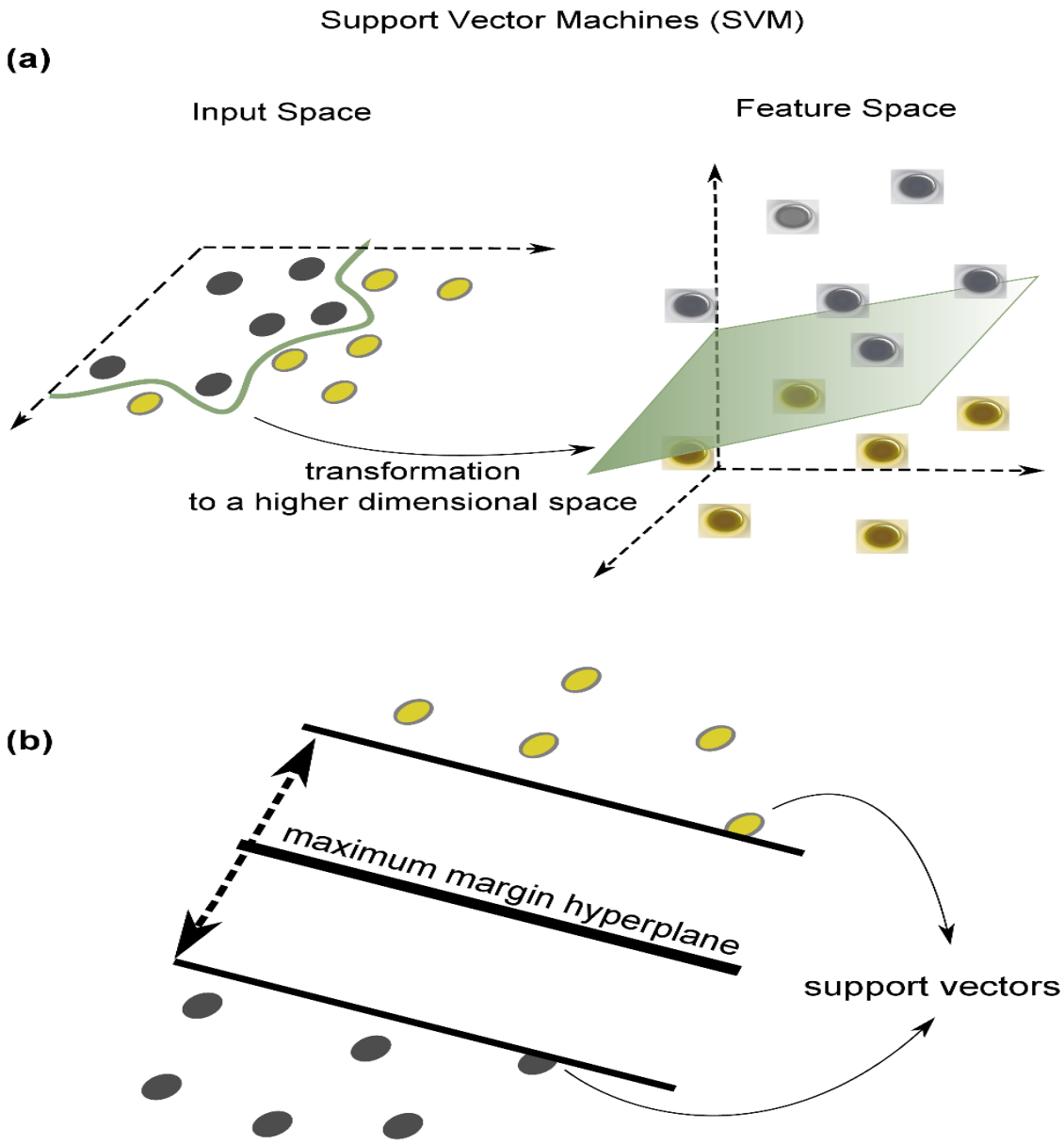


Figure 5.1: Illustrative example of the outcome of an SVM algorithm on a linearly separable binary problem. The optimal separating hyperplane maximizes the margin between the support vectors of each class.

origin. Therefore, the optimization task consists of finding the unknown parameters $\omega_k, k = 1, ..., S$ (S= number of voxels) by minimizing the norm of the vector $\mathbf{w}$ subject to some linear constraints defining the class belongings, that is:

$$arg_{min\mathbf{w},\omega_0\alpha\geq 0}\left\{\frac{1}{2}\parallel w \parallel^2 - \sum_{i=1}^{2}\alpha_i[y_i(\mathbf{w}.\mathbf{x}_i - b) - 1]\right\} \tag{5.3}$$

where α_i are some Lagrange multipliers in case of linear separable training data $\mathbf{x}_i$.

When no linear separation of the training data is possible, SVM can work effectively in combination with kernel techniques using the kernel trick, so that the hyperplane defining the SVM corresponds to a non-linear decision boundary in the input space that is mapped to a linearized higher-dimensional space [129].
In this sense, the decision function can be expressed in terms of the support vectors only:

$$f(\mathbf{x}) = \sum_{i=1}^{N_S}\alpha_i y_i F((\mathbf{s}_i, \mathbf{x}) + \omega_0) \tag{5.4}$$

where F($\mathbf{s}_i, \mathbf{x}$) is the kernel function, α_i are the solution for the optimization process, solved by either Quadratic Programming (QP) or the well known Sequential Minimal Optimization (SMO) [134], and $\mathbf{s}_i$ are the support vectors [129]. The chosen kernel function in different kinds of SVM, and the choice of the appropriate kernel for a specific application is a difficult task [88]. In this work, two different kernels were tested: the linear kernel and the radial basis function (RBF) kernel.

- The linear kernel function is defined as:
  $K(x_i, x_j) = 1 + x_i^T x_j$
- The RBF kernel is defined as:
  $K(x_i, x_j) = exp(-\gamma|x_i - x_j|^2)$

where γ is a hyperparameter (also called the kernel bandwidth).

5.2 Evaluation of classifier performance

5.2.1 Performance metrics

The purpose classifier consists to assign correctly a label to an object defined by feature vectors. For binary classification, in which data are divided into positive and

Table 5.1: Possible test results depending on the label.

		labels		
		Positive	Negative	
Test	Positive	TP	FP	→ *Positive predictive value*
	Negative	FN	TN	→ *Negative predictive value*
		↓	↓	
		Sensitivity	*Specificity*	

negative labels, there will be two possible errors can make the classifier: a pattern that is classified as a positive, however, it was actually a negative or vice versa.
These possibilities are reflected in table 5.1, where the test result is compared with the original label. Four possible test results can be obtained:

- True positive (TP) result when the two table parameters coincide positive value.
- False positive (FP) result when the test is positive while the original label was negative.
- False negative (FN) result when the test is negative while the original label was positive.
- True negative (TN) result when the two parameters coincide in negative value.

The sensitivity is defined as the ability of a classifier to detect the true positives and is expressed as:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5.5}$$

So that a sensitivity of 100% correspond to a classifier which is able to correctly classify all objects labeled as positive. Therefore, if a classifier with high sensitivity gives a negative result, it will be very reliable, which, in the case of computer aided diagnosis, can be used to discard the disease. The sensitivity is related to the error of type I in statistical inference, which is to reject the null hypothesis when it is actually true.
The specificity is defined as the ability of a classifier to detect the true negatives and is expressed as:

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5.6}$$

So that a specificity of 100% correspond to a classifier which is able to correctly classify all objects labeled as negative. In the case of computer aided diagnosis, a classifier with high specificity is very useful to confirm the disease. The specificity is related to the error of type II where the null hypothesis is accepted when it is actually false.
A good diagnostic test is one has both high sensitivity and specificity. The accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + TN + FP} \tag{5.7}$$

It is possible that a classifier has near of 100% and 0% of sensitivity and specificity values, respectively. This classifier does not have the capacity to discriminate between classes, as it would be a classifier to take any pattern as positive. This is equivalent to a random classification, since their accuracy will be around 50% for a sample without predominance of any of the two classes. The classifier is desirable that have a high value of accuracy, sensitivity and specificity simultaneously.
Other parameters that can be interesting are the predictive values. These parameters refer to the validity of a result of positive/negative classification (positive/negative predictive value). It can rely more on a classifier with high positive predictive value (PPV) than one with lower PPV. However, the predictive values depend on the preponderance of classes, called prevalence. This latter is a term of epidemiology which determines the proportion of individuals in a population that, in our case, have the disease. If the test set does not have an equal number of positive than negative, the positive or negative fractions (FPP/FPN) probability shall be used:

$$\text{FPP} = \frac{sensitivity}{1 - specificity} \tag{5.8}$$

and

$$\text{FPN} = \frac{1 - sensitivity}{specificity} \tag{5.9}$$

5.2.2 ROC curves

A receiver operating characteristic (ROC), or ROC curve, is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. ROC analysis has become a popular method for evaluating the accuracy of medical diagnostic systems. The curve is created by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) for different cut-off points. Each point on the ROC curve represents a sensitivity/specificity pair

corresponding to a particular decision threshold.
In figure 5.2, the ROC space is shown. The best possible classification method would yield a point in the upper left corner or coordinate (0, 1) of the ROC space, representing 100% of sensitivity value and 100% of specificity value. Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test [135].

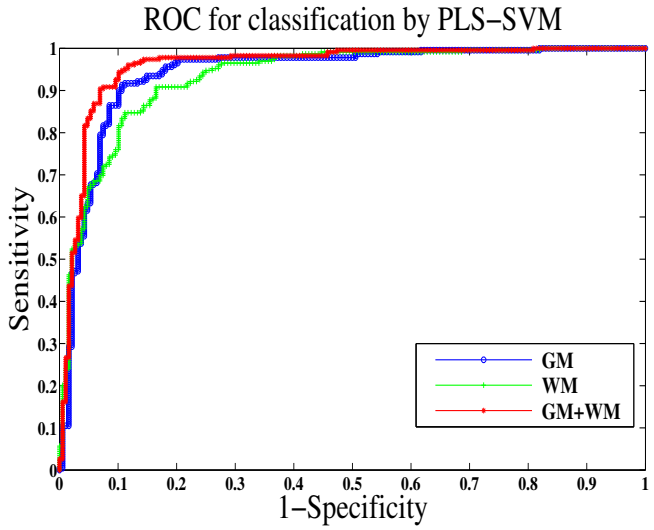


Figure 5.2: An example of a ROC Space. A ROC space is defined by (1 - specificity) and sensitivity as X and Y axes respectively, which depicts relative trade-offs between true positive rate and false positive rate.

5.3 Cross-validation methods

Cross-validation [136–138] is a validation technique used to assess how the obtained results of a statistical analysis will generalize to an independent dataset. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. A cross-validation technique is based on the segmentation of all the data into complementary subsets, and perform the analysis on a subset (called the training set) and validate this analysis in another subset (called testing set)[139, 140].
To reduce the variation of the result values during the global evaluation of the proposed system, multiple rounds of CV are performed using different partitions. In addition, the validation results are averaged over the rounds [141]. As a conclusion, CV method is based on averaging the measures of fit (prediction error) to derive a more accurate estimate of model prediction performance. In CV methods, the test set is not a "real" test, as the label elements set of test is known. In this sense, a comparison of the test result with the original label are performed to determine if it is

a TP, FP, TN, or FN. Once the process iterates over each partition, you can calculate any of the amounts defined in the previous section.

5.3.1 k-fold cross validation

k-fold cross validation is a common technique for estimating the performance of a classifier. In *k*-fold cross-validation, the data is firstly partitioned into k equally sized folds. Subsequently, k iterations of training and validation are performed such that within each iteration a different fold of the data is held-out for validation, while, the remaining $k-1$ folds are used for learning (see figure 5.3) [142]. Data is commonly stratified prior to being split into k folds. When, the stratification is the process of rearranging the data as to ensure each fold is a good representative of all information. For example, in a binary classification problem where each class comprises 50% of the data, it is best to arrange the data such that in every fold, each class comprises around half the instances.
The advantage of the *k*-fold cross validation method is that all the observations are used for both training and testing. Besides, each observation is used for validation only once. This method is often used when the number of elements in the sample is very large, or when sorting algorithms are computationally costs. In this case, it has control over the number of times that iterates through the validation number (*k*).

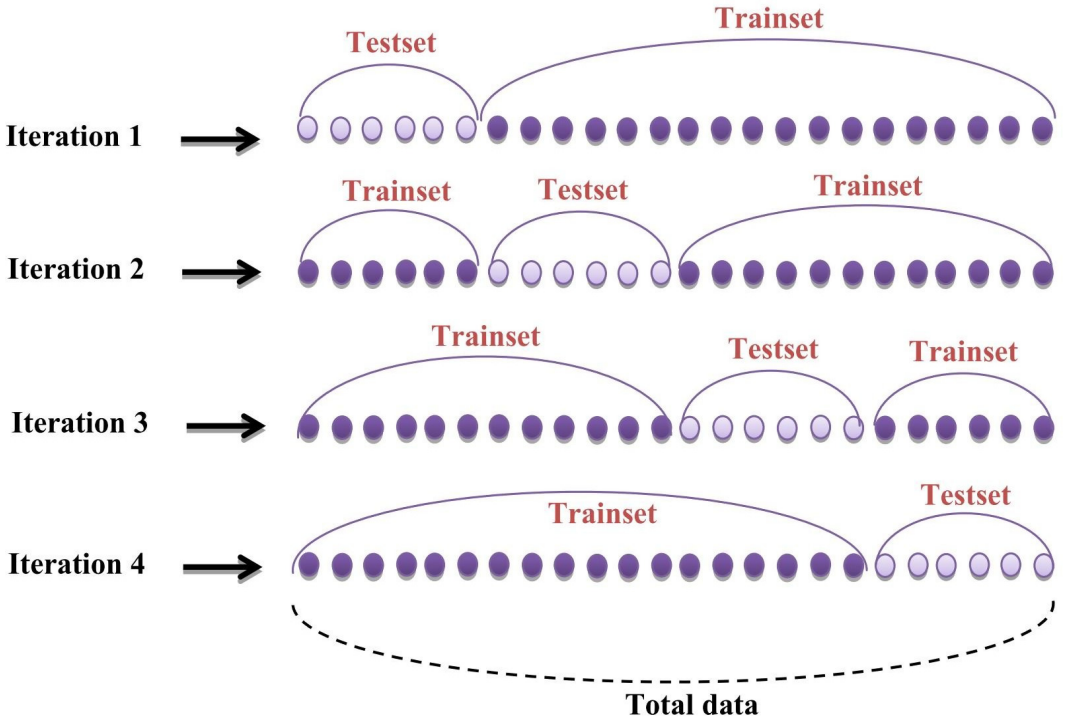


Figure 5.3: Scheme of *k*-fold cross validation.

5.3.2 Leave-one-out cross validation

Leave-one-out cross-validation (LOOCV) is a particular case of k-fold cross-validation where k equals the number of instances in the data. In other words, in each iteration nearly all the data except for a single observation are used for training and the model is tested on that single observation (see figure 5.4). An accuracy estimate obtained using LOOCV is known to be almost unbiased but it has high variance, leading to unreliable estimates [143]. It is still widely used when the available data are very rare, especially in bioinformatics where only dozens of data samples are available. The advantage of LOO CV method is that the training set is larger than the sample permits, increasing the statistical estimation of the parameters of the classifier. In contrast, this method can be computationally expensive given the large number of times that the validation process is repeated, if the number of elements in the sample is large.

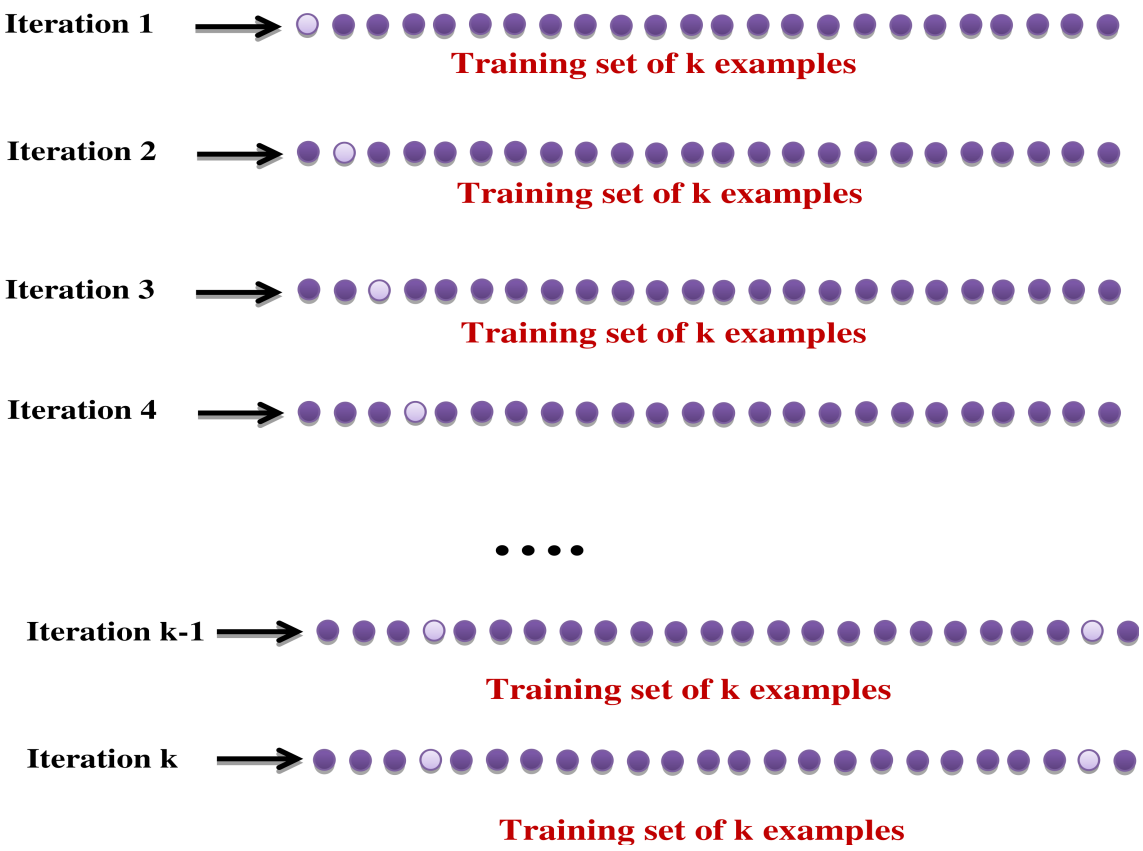


Figure 5.4: Leave-one-out cross-validation procedure.

5.4 Conclusion

This chapter has provided details of the machine learning method applied in this thesis work, namely, support vector machines. In addition, a description of their performance metrics may be assessed. Furthermore, the proposed CAD systems were carried out using the cross validation strategy which was also described in this chapter. The following four chapters present the main contributions of the thesis which detailed the results of the experimental development approaches. The two later chapters present respectively the conclusions of this work and a Spanish chapter which resume the work of this PhD thesis.

Part II

Experimental developments

Chapter 6

Structural MRI analysis based on Partial Least Square

PLS is a statistical model which is widely used as a method for modeling the relation between the observed variables by means of the latent variables. It has been a popular tool for dimensionality reduction techniques as well as the classification task. The underlying idea of all PLS methods is that the observed data are generated by a system which is running by a small number of latent variables [144].
PLS has received a great amount of attention of the scientific community due to its successful application in several areas such as bioinformatics, medicine, pharmacology, social sciences and physiology [106, 145–148]. A developed CAD system which is based on PLS approach have been proposed in this work for the early diagnosis of the disease. The mathematical details of this approach are presented in chapter 4 and the obtained results of the developed CAD system are presented in the following.

6.1 Experiments

A combination between the PLS feature extraction approach and the SVM classifier have been evaluated for different brain tissues from the segmented sMRI brain images. All the experiments are carried out by the k-fold cross-validation strategy, that is, the complete classification system is trained by taking into account all the samples but one, which is used as test samples. This procedure is repeated as many times as samples in the database, leaving each sample out in each iteration. Finally, an average accuracy rate is computed. k-fold has been used to assess the discriminative accuracy of the proposed CAD system applied to the discrimination of frontotemporal dementia from AD [149] and in classifying atrophy patterns based on sMRI images [150].
This experimental section is aimed to achieve two objectives. Firstly, by reducing the dimensionality of the input images in order to solve the sample size problem. The proposed methodology based on t-test feature selection and PLS feature extraction approaches demonstrates its effectiveness for segmented sMRI images modeling without loss of relevant information. Secondly, it addresses to the ability of the proposed CAD system which provides good results of accuracy, sensitivity and specificity rates.
Figure 6.1 illustrates the pipeline of the proposed CAD system.

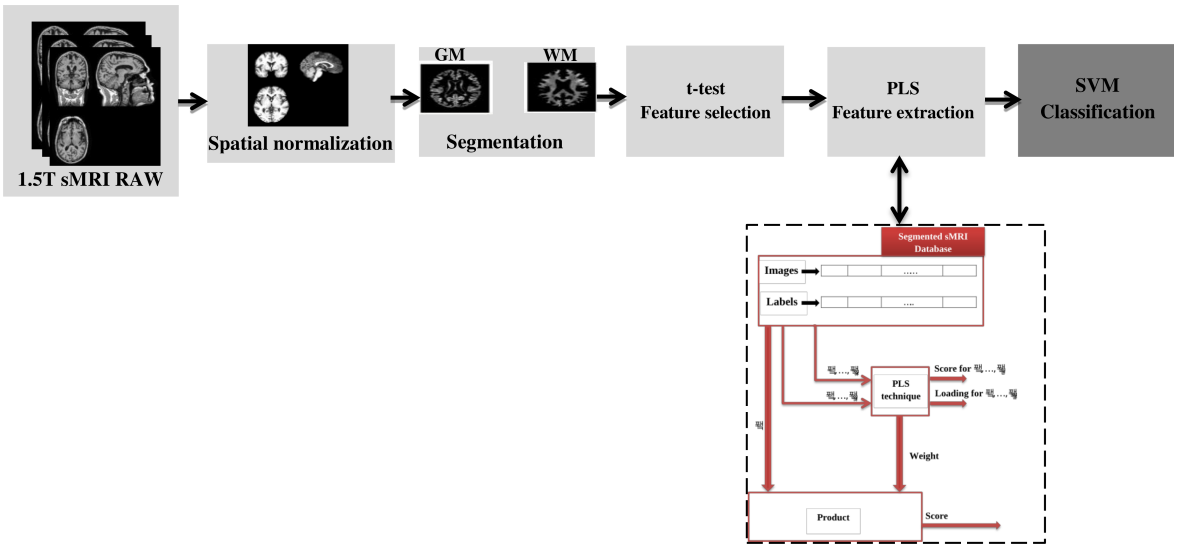


Figure 6.1: Schema of the proposed PLS-CAD system.

For the analysis of the proposed CAD system, the ADNI database that is used in this PhD thesis, is arranged into three different groups:

- **Group 1**: only AD (positive) and Normal controls (negative) patient images are considered.

- **Group 2**: only MCI (positive) and Normal controls (negative) patient images are considered.

- **Group 3**: only MCI (positive) and AD (negative) patient images are considered.

6.2 Results

The classification results are summarized in table 6.1 with different experiments (GM images, WM images and the combination of feature extracted from GM and WM brain tissues), using PLS feature extraction approach and different SVM classifiers.

Table 6.1: Statistical measures of performance of PLS feature selection method with different SVM classifiers, for the three sample groups, and using eight components.

		Group 1	Group 2	Group 3
Brain tissues	Kernel	Acc/Sens/Spec(%)	Acc/Sens/Spec(%)	Acc/Sens/Spec(%)
GM	Linear	**87.53/88.65/86.17**	**77.57/76.76/71.9**	**77.03/74.59/79.46**
	RBF	87.29/87.77/86.7	76.22/81.62/70.8	76.22/74.59/77.84
WM	Linear	**85.61/87.34/83.5**	**80.54/79.46/81.62**	**87.03/88.65/85.41**
	RBF	84.41/85.59/83	81.35/76.22/80.54	86/85.41/86.49
GM+WM	Linear	**88.49/91.27/85.11**	**81.89/82.16/81.62**	**85.41/87.03/83.78**
	RBF	88.49/90.39/86.17	80.27/73.51/82.7	85.41/85.95/84.86

- **Classification result of group 1**

PLS + Linear SVM method yielded higher accuracy rates than RBF. Furthermore, the combination of features extracted from GM and WM segmentation reported a classification accuracy of 88.49% for PLS and linear SVM (sensitivity= 91.27% and specificity= 85.11%) compared to 87.53% for GM only (sensitivity= 88.65% and specificity= 86.17%) and 85.61% for WM only (sensitivity= 87.34% and specificity= 83.51%). As a conclusion, the combination between features extracted from both GM and WM tissue distributions increases the classification rates of the classifier.

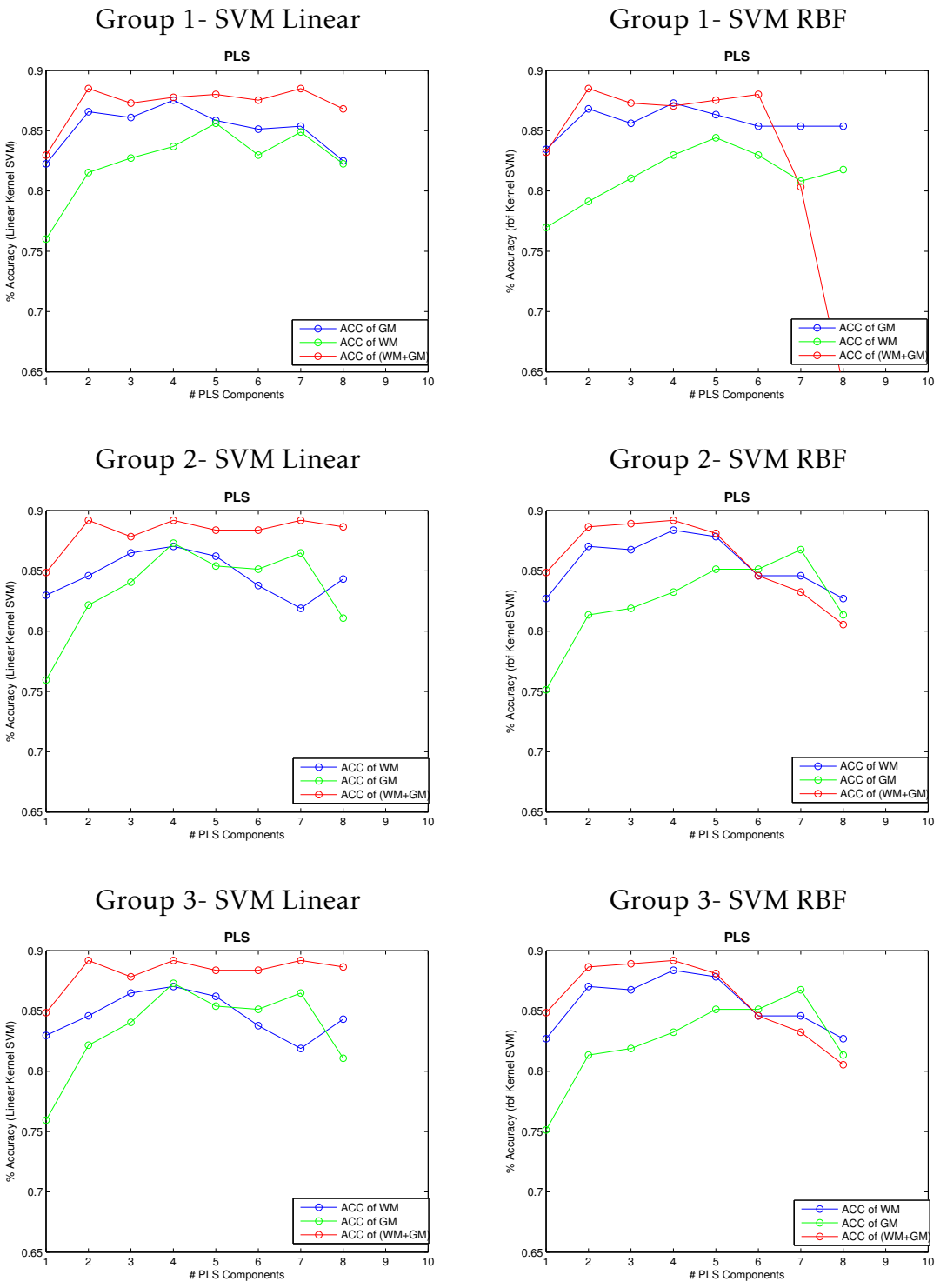


Figure 6.2: SVM classification: Values of Accuracy (%) computed for ADNI database in function of number of component for PLS feature extraction technique: with SVM linear (left) and RBF (right): (a) and(b) are the results of group 1, (c) and (d) are the results of group 2, (e) and (f) are the results of group 3.

- **Classification result of group 2**

The most difficult classification task concerning the ADNI database is to distinguish between NC and MCI patients, due to the wide range spanned by the features extracted from MCI patients. Using PLS and linear SVM, the combination of features extracted from GM and WM brain tissues provided the highest accuracy, 81.89% (sensitivity=82.16% and specificity=81.62%), whereas the features extracted from GM or WM brain tissues alone reported a classification accuracy of 77.57% and 80.54% respectively. Overall, we note that combining features extracted from both GM and WM brain tissues yielded the highest accuracy value.

- **Classification result of group 3**

Combining features extracted from GM and WM segmentation reported a classification accuracy of 85.41% for PLS and linear SVM (sensitivity= 87.03% and specificity= 83.78%) compared to 77.03% for GM only (sensitivity= 74.59% and specificity= 79.46%) and 87.03% for WM only (sensitivity= 88.65% and specificity= 85.41%). These results showed that the most important change in the brain occurs more in the white matter than in the gray matter brain tissues [151].
As shown in the table 6.1, the proposed and the analyzed method highlights that the combination of features extracted from GM and WM brain tissues distributions give better accuracy, sensitivity and specificity values than using different brain tissues separately. As a result, combining the different features extracted from both brain tissues (GM+WM) of patients with classification methods produces a valid approach to perform a CAD system for AD.
The accuracy value of the proposed model depends on the size of the feature vector. Therefore, the maximum size of the feature vectors is the number of brain image database minus two. However, this number may be reduced by selecting only the most important components. These components contain the most relevant information related to AD.
Figure 6.2 shows the accuracy rates of the different groups achieved with PLS approach in function of number of components selected (number of components= 8).

6.3 Discussion

The proposed method which is analyzed in this chapter is valid approach to develop CAD system for AD. In addition, this methodology achieves good values of accuracy, sensitivity and specificity.
The successful rate of PLS based method reached 88.49% for group 1. However, it is decreased for group 2 and 3 (78.92% ,85.4% respectively) when MCI images are

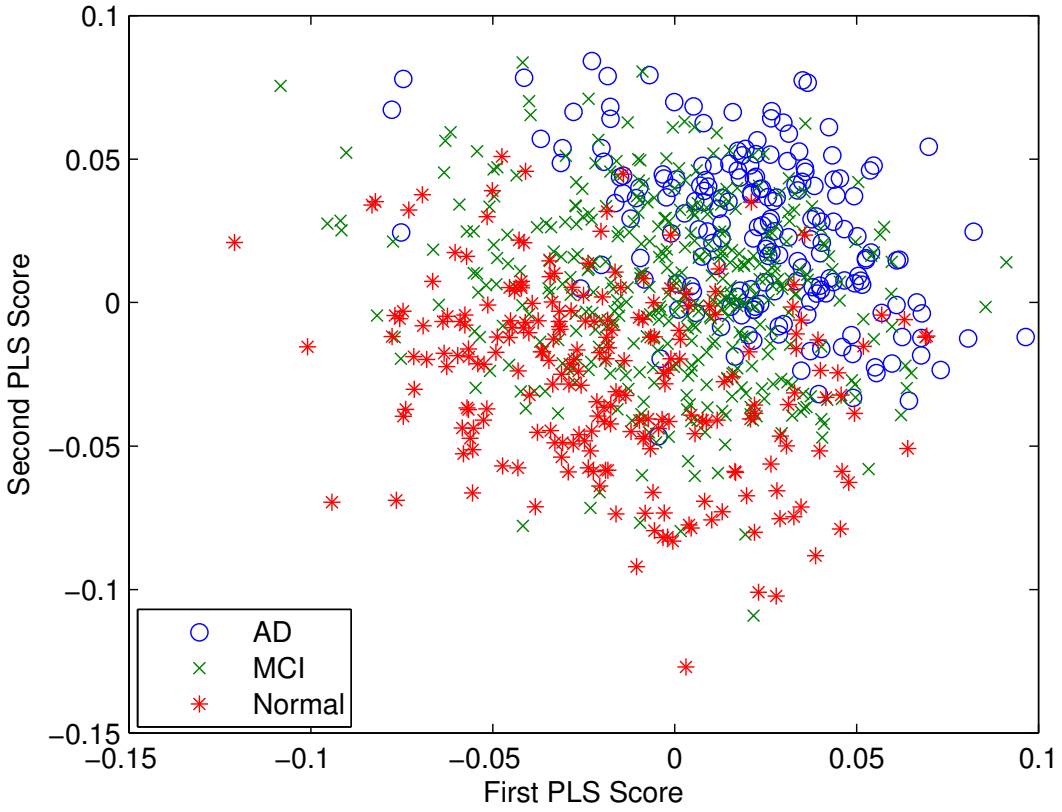
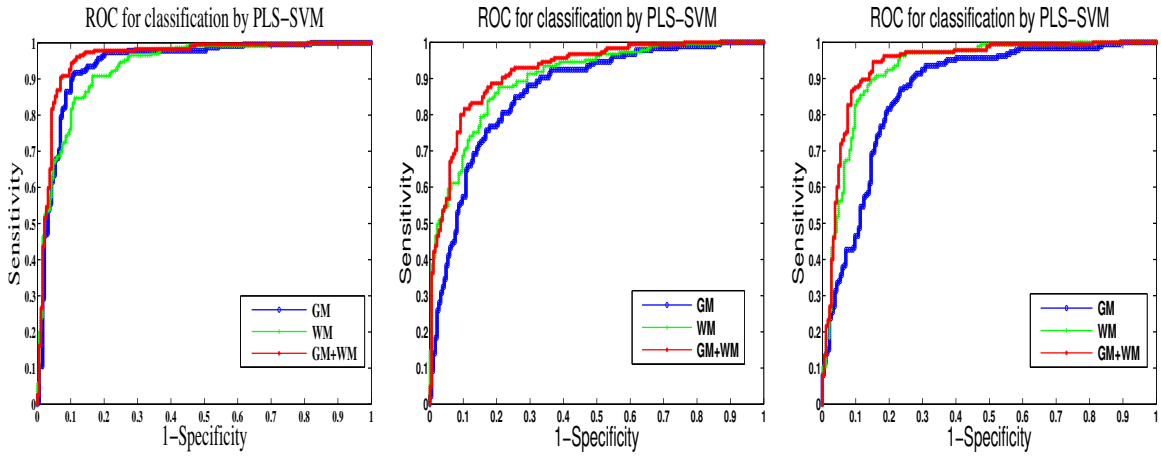


Figure 6.3: Representation of all images of the database using only the two first Scores of PLS.

included (see figure 6.2). This is probably due to the high variability of the MCI pattern of each image. As a consequence, the classification task becomes more difficult (see figure 6.3). As shown in the result of group 2 (NOR. vs. MCI), the classification result using only WM brain tissue is better than using only GM. This result confirms the previous study [152] in which the modification in the pattern of brain atrophy in early study (MCI) of disease occurs in the WM brain tissue. Besides, elder subjects are likely to have WM structural abnormalities caused by leucoaraiosis or other diseases [153]. This abnormality in the WM brain tissue for patients with AD or MCI can make the structure very different from normal controls. Thus, the classification result in the WM can be better than in GM of brain images. Furthermore, the classification results for group 3 (MCI. vs. AD) confirms that neurodegeneration starts in the WM and spreads to GM with the progression of the disorder.
It is worth noting that CAD systems are reproducing current medical knowledge since they have been trained with samples labeled by physicians. For this reason, statistical measures reported in this chapter are an estimation about how a trained system is able to reproduce a medical diagnosis performed by experts [154]. Thus,

some possible errors in the labeling process can modify the decision hyperplanes of the classifiers, specifically considering that the labels were assigned based on the scores obtained by patients in cognitive tests (as MMSE and CDR).
In this work, we have previously shown that combining features extracted from GM and WM segmentation gives a good classification accuracy using PLS approach. In addition, PLS + Linear SVM model yields good classification results with smaller computational time.
A more interesting approach consists of selecting only some components of feature extraction methods, such as FDR as it is described in [155] and the Out-Of-Bag (OOB) error in [154]. This previous methods achieve that using the first PLS components is optimal for classification purposes. In the development of our CAD system, we have selected only the 8 first PLS components. A higher number of components may worsen the classification results since it increases the input space. The relevance of the classification results obtained is also confirmed by the receiver operating characteristic (ROC) curves shown in figure 6.4. This figure gives us more information about the SVM classifier performances: the SVM with linear kernel reached the best trade-off between sensitivity and specificity. In addition, the closer to the left upper corner values are the better [156] which is the case of the combination of brain tissues in this work.



(a) NOR. vs. AD (b) NOR. vs. MCI (c) MCI. vs. AD

Figure 6.4: Comparison of the trade off between sensitivity and specificity. ROC curves for the three cases analyzed: using only GM images, using only WM images and using both GM images and WM images.

6.4 Conclusion

In this chapter, a complete CAD system for the diagnosis of the early AD has been presented. The proposed CAD system combines a feature extraction and a feature selection approaches with classification technique to improve the analysis of the segmented sMRI database to diagnosing AD. The proposed methodology based on two approaches which allow to reduce the dimension of the feature vector in order to surmount the small-size problem which arises in classification problems when the dimension of the feature vector is very high compared to the number of available samples.
The first approach uses t-test feature selection approach in order to extract the most discriminant regions from the different brain tissues of the segmented sMRI images. The second approach uses score vectors obtained through a Partial Least Squares algorithm as features. Score vectors are chosen following a criterion of maximum covariance between images and labels.
The proposed methodology has been tested, using two SVM kernels, the linear and the nonlinear (RBF) kernels. The resulting CAD system was trained using sMRI images from the ADNI database and the statistical performances of the novel aided diagnosis system were estimated using a k-fold cross-validation methodology.

Chapter 7

Structural MRI analysis based on Principal Component Analysis

PCA [157] is a standard technique to extract the most significant features from the brain image. Moreover, PCA is based on the action of a linear transformation (also known as the Karhunen-Loeve transformation) on a dataset of zero mean that diagonalized the covariance matrix. Mathematically, PCA is defined as an orthogonal linear transformation that transforms the initial correlated variables in a small number of uncorrelated variables, called Principal Components (PC). The first principal component will contain the characteristics of the data that have a greater contribution to the variance, followed in descending order by its value of variance for the second principal component, the third component, etc.
The PCA approach has been applied successfully in different image classification problems [91, 158, 159], and in particular in neuroimaging classification [155]. An automatic CAD system which is based on PCA approach has been developed in this work. The mathematical details of this approach are presented in chapter 4. However, the results of this developed CAD system will be presented in detail in this chapter.

7.1 Experiments

The feature extraction technique based on the PCA approach in combination with SVM classifier has been tested on the segmented sMRI images from the ADNI database. Then, a binary mask for each tissue brain images (GM images and WM images) is computed by averaging all the normal subject tissue images. Only the voxels that have an intensity above 10% of maximum intensity in the average image will be considered. Thus, the application of a binary mask reduced the voxel numbers; for $121 \times 145 \times 121$ voxels (2,122,945 voxels) to 382,325 voxels. After that, a t-test feature selection approach is applied to select the most relevant features. Then, the PCA approach is applied to reduce more the input space and to extract the principal component vectors which are used as input for a statistical classifier.

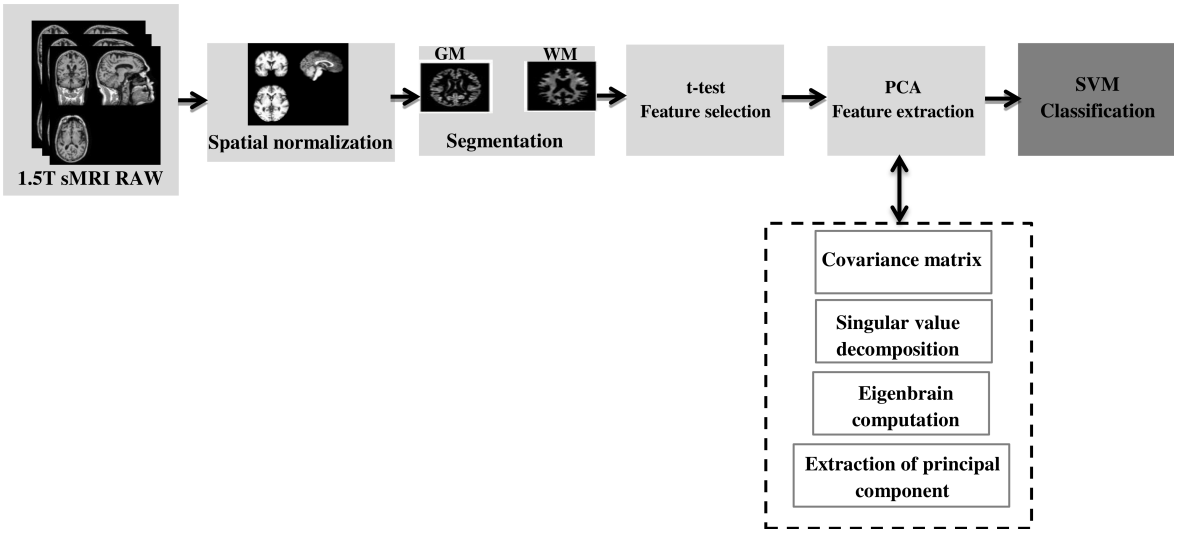


Figure 7.1: Schema of the proposed PCA-CAD system.

The main objective of these experiments is to determine at first the ability of PCA analysis as a dimension reduction technique of the space and to determine discrimination features to distinguish between subjects labeled as normal, MCI or AD patients. Once the features are extracted and reordered according to the most convenient criterion (variance), it is intended to see how they are in accordance with different classifiers, evaluating their performance by varying parameters inherent in them. The maximum number of PCA coefficients that can be extracted from a database of N samples is N-1. The experiments were performed by varying the number of coefficients m used in training and testing tasks from m =1 to m = N-1, to determine in each case the appropriate number of coefficients is necessary for successful separation of classes. On the other hand, SVM classifiers with linear and RBF kernels type are evaluated.

The description of the groups on which all the process is applied are:

- **Group 1**: 360 subjects; 185 as NORMAL and 185 as AD.
- **Group 2**: 360 subjects; 185 as NORMAL and 185 as MCI.
- **Group 3**: 360 subjects; 185 as MCI and 185 as AD.

7.2 Results

Table 7.1 shows the statistical measures obtained using PCA approach. The performance of this feature extraction technique was calculated by means of k-fold cross-validation with a number of folds equal to 10 (k=10). Thus, this table presents the values of the accuracy, sensitivity and specificity for the different brain tissues (GM, WM, GM+WM) using the proposed PCA approach and the different SVM classifiers (with linear and non linear kernels).

Table 7.1: Statistical measures of performance of PCA feature selection method with different SVM classifiers, for the three sample groups, and using eight components.

		Group 1	Group 2	Group 3
Brain tissues	Kernel	Acc/Sens/Spec(%)	Acc/Sens/Spec(%)	Acc/Sens/Spec(%)
GM	Linear	**85.61/89.08/81.38**	**75.41/78.38/72.43**	**72.71/72.43/72.97**
	RBF	83.93/86.26/81.38	72.97/75.68/70.27	71.08/71.89/70.27
WM	Linear	**81.77/84.28/78.72**	**75.14/77.3/72.98**	**79.19/82.16/76.22**
	RBF	81.29/83.41/78.72	72.7/64.86/80.54	74.86/74.05/75.68
GM+WM	Linear	**87.77/89.96/85.11**	**78.92/80/77.84**	**81.89/84.86/78.92**
	RBF	87.55/90.39/84.04	73.24/71.89/74.59	76.77/74.59/78.92

- **Classification results for group1**

The classification results obtained using *group1*, which consisted on distinguishing between NC and AD subjects, show a higher value of the accuracy rate.

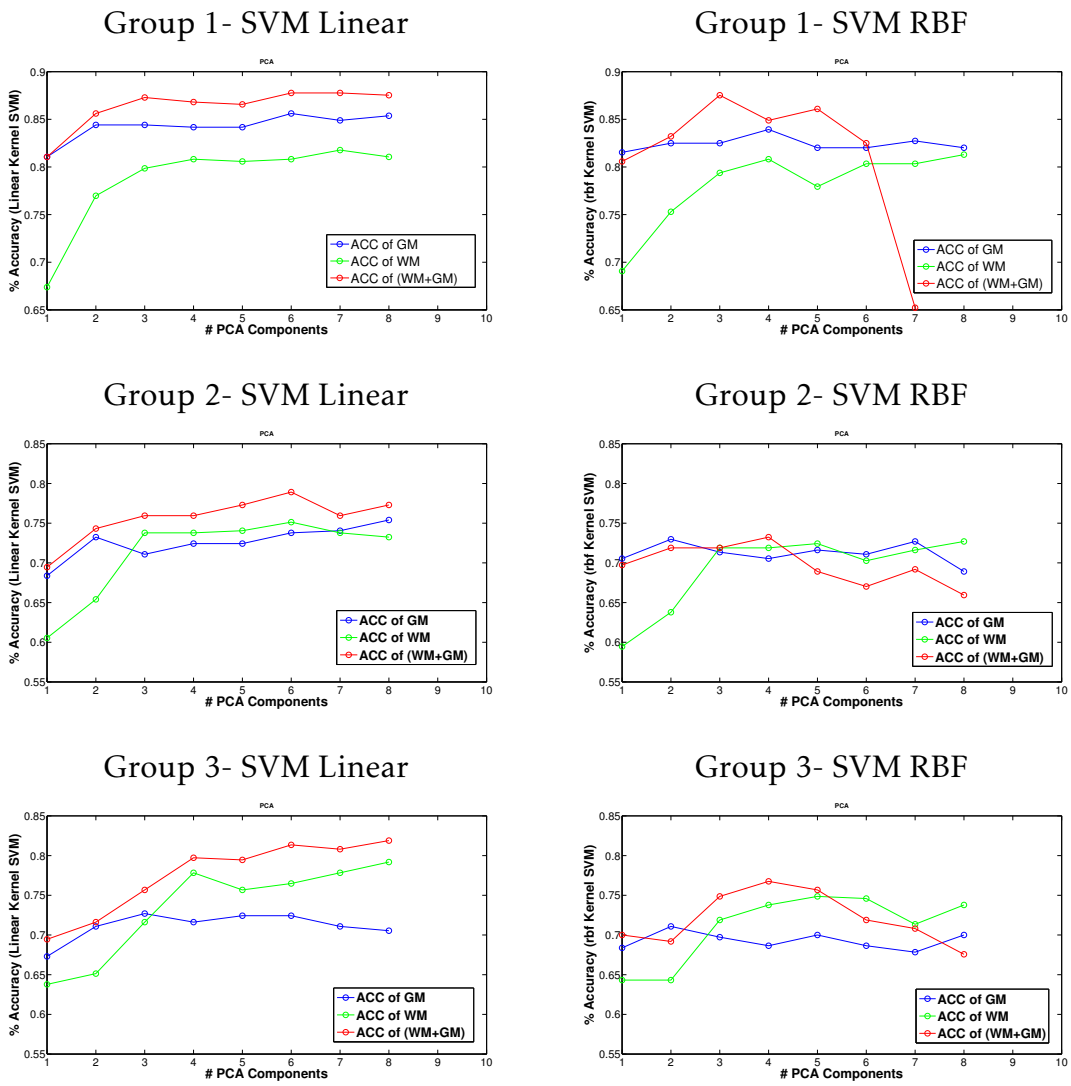


Figure 7.2: SVM classification: values of Accuracy (%) computed for ADNI database in function of number of component for PCA feature extraction technique: with SVM linear (left) and RBF (right): (a) and (b) are the results of group 1, (c) and (d) are the results of group 2, (e) and (f) are the results of group 3.

Figures 7.2.(a) and 7.2.(b) show that the linear-SVM yielded a higher accuracy rates than RBF. Combining features extracted from GM and WM segmentation reported a classification accuracy of 87.77% for PCA and linear SVM (sensitivity of 89.96% and specificity of 85.11%) compared to 85.61% for GM only (sensitivity of 89.08% and specificity of 81.38%) and 81.77% for WM only (sensitivity of 84.28% and specificity of 78.72%). As a conclusion, combining features extracted from both GM and WM tissue distributions increases the classification and accuracy of the classifier.

- **Classification results for group2**

The most difficult classification task in this work is to distinguish between NC subjects and subject in the early asymptotes of the disease (MCI subjects), due to the wide range spanned by the features extracted from MCI. Using PCA and linear SVM, the combination of features extracted from GM and WM segmentation provided the highest accuracy, 78.92% (sensitivity 80% and specificity 77.84%), whereas the features extracted from GM or WM alone reported a classification accuracy of 75.41% and 75.14% respectively. Overall, we note that combining features extracted from both GM and WM tissue yielded the highest accuracy value. The classification results related to this experiment are shown in figures 7.2.(c) and 7.2.(d).

- **Classification results for group3**

Table 7.1, figures 7.2.(e) and 7.2.(f) show the classification results obtained in the last experiment, which consisted on distinguishing between MCI and AD subjects using different SVM classifiers. The Combination of features extracted from GM and WM segmentation are given a classification accuracy value of 81.89% (sensitivity value of 84.86% and specificity value of 78.92%) compared to 72.71% for only GM tissue brain images (sensitivity value 72.43% and specificity value of 72.97%) and 79.19% for WM only (sensitivity 82.16% and specificity of 76.22%). As shown in the table 7.1, the PCA method analyzed in this work highlight that the combination of features extracted from GM and WM tissue distributions gives better accuracy, sensitivity and specificity than using different brain tissues separately. As a result, combining the different features extracted from both brain tissues (GM and WM) of patients with classification methods produces a valid approach to perform a CAD system for AD.

7.3 Discussion

As shown in the above section that PCA is a valid approach to develop a CAD system for AD. In addition, the developed CAD system achieve a good value of accuracy,

sensitivity and specificity. From figure 7.2, we can note that the performance of the CAD system improves with the number of PCA components used as input features for classification which up to a maximum stable value.
The successful rate of PCA based method reached 87.77% for group1. However, it is decreased for group 2 and 3 (78.92% and 81.89% respectively) when MCI images are included (see figure 7.2.(c) and 7.2.(d)). This is probably due to the high variability of the MCI pattern of each image. As a consequence, the classification task becomes more difficult (see figure 7.3).

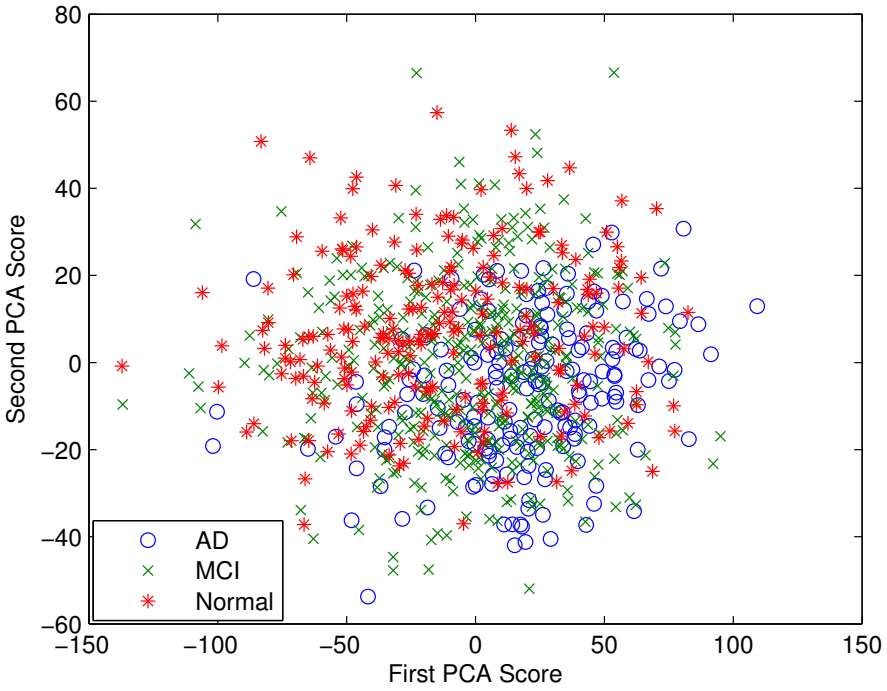
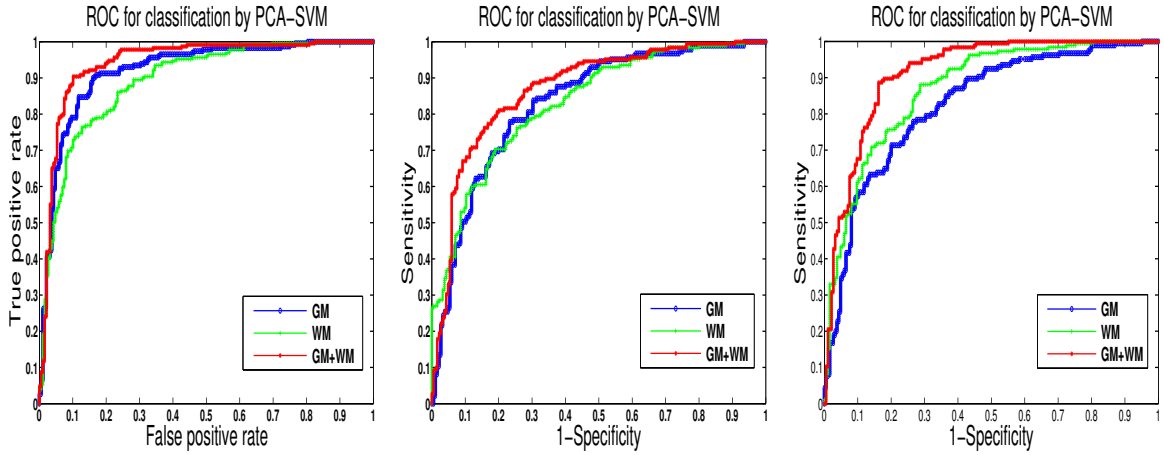


Figure 7.3: Representation of all images of the database using only the two first Scores of PCA.

In order to evaluate the proposed CAD system, a receiver operating characteristic (ROC) curves are used.
Figure 7.4 contains the sensitivity and 1-specificity values of the classification results in the ROC space considering the PCA technique. These plots show a trade-off between the specificity and sensitivity of the CAD system when varying any of the input parameters. In addition, the closer to the left upper corner values are the better [156] (the combination of brain tissues, i.e. GM and WM).



(a) NOR. vs. AD (b) NOR. vs. MCI (c) MCI. vs. AD

Figure 7.4: Comparison of the trade off between sensitivity and specificity. ROC curves for the three cases analyzed: using only GM images, using only WM images and using both GM images and WM images.

7.4 Conclusion

A CAD system for assisting the early detection of the Alzheimer's disease was shown in this chapter. The system was developed by combining the different brain tissues and using the t-test and PCA as feature selection and extraction methods. Thus, these approaches allow the dimensionality reduction of the feature vector in order to surmount the small sample size problem. The classification task is performed using an SVM classifier with linear and RBF kernels, and performance values are obtained via k-fold cross-validation. The proposed system shows good classification rates in the discrimination task between AD and NC images. Moreover, it yields a peak value up to 88% for the accuracy, 90% for the sensitivity and 85% for the specificity, when combining features extracted from GM and WM brain tissues, and using linear SVM. When MCI brain images are included in the classification, the accuracy value is slightly affected (79% and 82% for NC. vs. MCI and MCI. vs. AD, respectively). As a conclusion, these obtained results reveal that the developed system, improve the classification rates of sMRI ima-ges for an accurate diagnosis of the early AD stage.

Chapter 8

Structural MRI analysis based on Independent Component Analysis

The statistical independence is a desirable property in the field of pattern recognition, and Independent Component Analysis (ICA) approach provides the necessary to obtain a set of independent sources from a data set. In signal and image processing field, the usual framework in which ICA arises is the problem of "cocktail party effect". Thus, its solution is to recover the information from the original sources that produced a recorded signal mixture from the Blind Source Separation (BSS) process. The resolution of this problem must be carried out without any knowledge about the proportion source / noise generated by the recorded signal. Through ICA, seeking the statistical independence of the original sources, you can obtain the set of original sources, obtaining with ICA several advantages over other methods that can solve the problem, as PCA. One of these advantages is that the noise sources can be completely separated and removed with ICA approach, so that a separate set of sources containing all the relevant information is obtained. A developed CAD system based on ICA have been proposed in this work to allow us to extract highly representative features, which are closely related to typical AD patterns, and to improve the classification results in the early stage of AD.

8.1 Experiments

Two different methodologies were realized with ICA approach and segmented sMRI brain images:

- **Method I**: is based on a combination between t-test and ICA, as a feature selection and extraction approaches.
- **Method II**: is based only on ICA feature extraction approach from the class mean images.

For testing early diagnosis capabilities, MCI was considered as a pre-stage of AD [150, 160, 161] and the following three groups were designed:

- **Group 1**: Only NC subjects and patients with probable AD are considered and is referred to as "NC. vs. AD" in the following.
- **Group 2**: Only MCI and NC patient images are considered and is referred to as "NC. vs. MCI".
- **Group 3**: Only MCI and AD patient images are considered and is referred to as "MCI. vs. AD".

8.1.1 Method I

8.1.1.1 Application to the segmented sMRI database

In this work, we have a matrix X that contains all x_n mixture vectors, each one containing N values. Each vector represents the intensity values for each brain image. An application of the t-test feature selection approach to the matrix X established to reduce the dimensionality by selecting the first 8000 relevant feature voxels. After that, an application of the FastICA algorithm is established to the new reduced matrix for extracting the first K eigenvalues (k=8; the number of independent components). Then, the independent components are extracted using the algorithm described in chapter 4. From this algorithm, we obtain the mixing matrix A and its inverse W for the complete database X. After that, classification experiments were performed to distinguish between NC, MCI and AD. The pipeline of this proposed methodology presents in the figure 8.1.

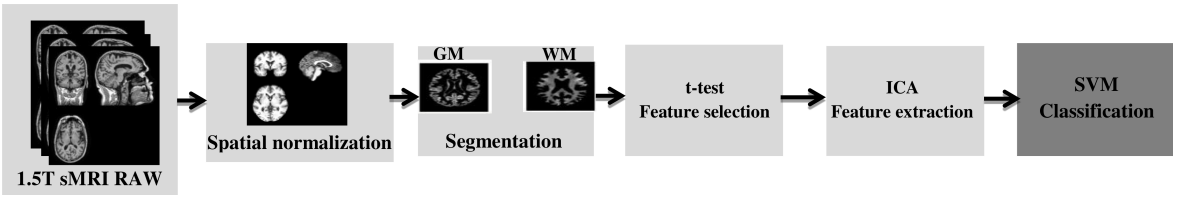


Figure 8.1: Illustration of the system used in "method I".

8.1.1.2 Results and Discussion

The classification results are summarized in table 8.1 with different experiments (GM images, WM images and the combination of feature extracted from GM and WM brain tissues), using ICA feature extraction approach and different SVM classifiers. It can be noted from this table that the highest accuracy value is obtained for the group 1 and it is decreased in group 2 and 3 when MCI images are included; this diminution of their performance metrics can be explained by the high variability of the MCI pattern of each image. Second, it can be concluded from the measures of performance metrics that the linear SVM classifier yields a higher accuracy rate with ICA feature extraction approach than SVM-RBF (see figure 8.2). Thus, the linear SVM might be the best technique to distinguish AD and MCI patients from NC. Third, it is remarkable that when using the combination of feature extracted from GM and WM images as input features and afterwards transformed them with ICA coupled with linear SVM increase the accuracy of the classifier. Thus, adds a valuable robustness to our system for distinguishing AD and MCI from NC subjects.

Table 8.1: Statistical measures of performance of ICA feature selection method with different SVM classifiers, for the three sample groups, and using eight components.

		Group 1	Group 2	Group 3
Brain tissues	Kernel	Acc/Sens/Spec(%)	Acc/Sens/Spec(%)	Acc/Sens/Spec(%)
GM	Linear	**84.65/86.46/82.45**	**69.46/69.03/69.96**	**69.19/70.27/68.11**
	RBF	82.97/83.41/82.45	68.38/68.65/68.11	64.05/61.62/66.49
WM	Linear	**70.26/72.93/67.02**	**63.51/63.24/63.78**	**59.46/62.16/56.76**
	RBF	68.82/68.56/69.15	61.08/56.22/65.95	53.51/51.89/55.14
GM+WM	Linear	**86.37/88.34/83.98**	**70.19/72.89/67.49**	**69.83/73.43/66.24**
	RBF	78.66/75.11/82.98	65.13/61.08/69.19	64.32/51.89/76.76

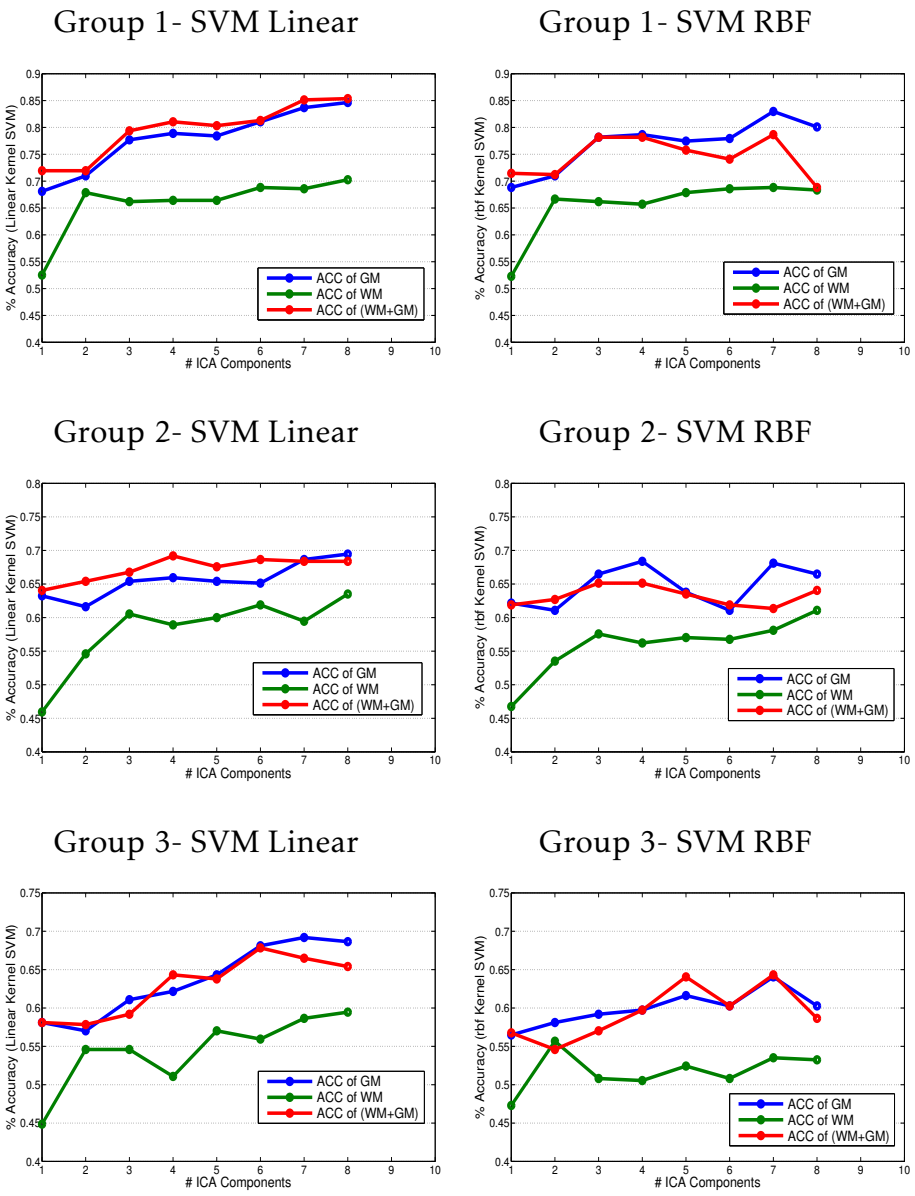


Figure 8.2: SVM classification: Values of Accuracy (%) computed for ADNI database in function of number of component with ICA feature extraction: (Above) the classification accuracy of group 1 (NOR versus AD), in (the middle) the classification accuracy of group 2 (NC versus MCI) and (below) the results of group 3 (MCI versus AD).

8.1.2 Method II

In this section, a CAD system is proposed with special emphasis in the whole structure of the system, that is, not only relying on performance measures but also on diagnosis support. Several pattern recognition and machine learning techniques are combined to provide an accurate decision and a visual support.
Figure 8.3 gives an overview of the procedure used in this CAD system. The image under analysis is decomposed into its class relevant patterns and a two-dimensional (2D) scatter plot is used to represent the degree of class belongings according to those patterns. To obtain class representative patterns, the sMRI segmented into WM and GM tissues brain is used to grow 2D subspaces. The basis that spans these subspaces, talking in algebraic terms, is made up of a basis of independent vectors obtained by the ICA on means (ICAm) method. The images are orthogonally projected onto this subspace and classified using the SVM classifier, furnishing the classification output with a graphical interpretation.

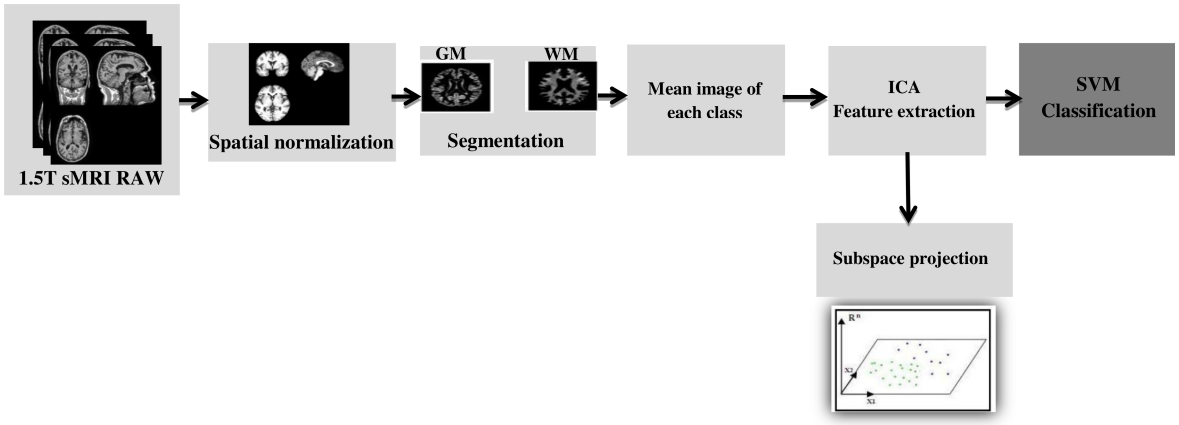


Figure 8.3: Flow diagram of the procedure used in the proposed CAD system.

8.1.2.1 Projection onto representative subspaces

Consider a n dimensional vector space E in which each sMRI image of the brain is a vector $\mathbf{v}$. Additionally, each vector $\mathbf{v}$ has a corresponding label c, which typically can only take two values $\{c_1, c_2\}$, forming a pair $\{\mathbf{v}, c\}$. The coordinates of the vector v_i, expressed in the standard orthonormal basis $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ on the space E, correspond with the values the image takes on each voxel, ranging from 1 to n, the total number of voxels in the image:

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + ... + v_n\mathbf{e}_n \tag{8.1}$$

The idea behind the projection onto representative subspaces is to find a subspace

$S \subset E$ with dimension $m \ll n$ spanned by a basis of vectors $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m\}$ and orthogonally project each image vector onto this subspace. The vectors $\mathbf{u}_i$ are the images that contain the most representative patterns of each class c_i, whose extraction process is described below. In addition, the dimension m will correspond with the number of classes. Therefore, each image projection onto S is expressed as:

$$\mathbf{v}_{\|} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + ... + x_m\mathbf{u}_m \tag{8.2}$$

It is assumed that the orthogonal complement of the subspace S, that is $S^{\perp}$, contains the information related to all the other variations in size and form and not related to the classes. Therefore, this projection is considered as a feature extraction method, and the resulting feature vector is made up of the coordinates x_i. Since, the vectors $\mathbf{u}_i$ are not necessarily orthogonal, to calculate x_i one needs to make use of the projector operator P_A: Let $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ be a basis of the subspace S, and let A denote the n-by-m matrix whose columns are $\mathbf{u}_1, ..., \mathbf{u}_m$. Then, the projection is given by: [162]

$$P_A = A(A^T A)^{-1} A^{\mathrm{T}} \tag{8.3}$$

so that the application of that operator on a vector image:

$$\mathbf{v}_{\|} = P_A\mathbf{v} \tag{8.4}$$

gives the coordinates of equation 8.2. The procedure to determine the basis $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ of the representing subspace is therefore crucial to successfully capture the class-relevant information.

8.1.2.2 Extraction of a representative vector basis by ICA means (ICAm)

Figure 8.5 shows brain image patterns that represent the ICs of an image decomposition; two vectors that span with a class-representative subspace. To obtain them one should first compute the class mean images:

$$\bar{\mathbf{v}}_i = \frac{1}{N_m} \sum_{j=1}^{N_m} \eta_{ij}\mathbf{v}_j, \qquad i = 1, 2, ..., m. \tag{8.5}$$

where N_m is the number of vectors belonging to the class c_i and:

$$\eta_{ij} = \begin{cases} 1 & if\ \mathbf{v}_j \in c_i \\ 0 & otherwise \end{cases} \quad i = 1, 2, ..., m \tag{8.6}$$

If $\mathbf{X}$ is a n-by-m matrix, whose columns are the $\bar{\mathbf{v}}_1, ...\bar{\mathbf{v}}_m$ vectors, then ICA can be applied to obtain a set of independent vectors. Let $\mathbf{X}$ be an observed random vector and $\mathbf{A}$ is a full rank matrix such that:

$$\mathbf{X} = \mathbf{AS} \tag{8.7}$$

where the source signals $\mathbf{S}$ contain the independent components: $p_s(\mathbf{S}_1, ..., \mathbf{S}_n) = p_{s1}(\mathbf{S}_1)...p_{s_n}(\mathbf{S}_n)$. Thus, ICA recovers both the sources $\mathbf{S}_j$ and the mixing process using the independence assumption.
In the linear case, this latter task consists of finding the mixing matrix $\mathbf{A}$. A popular approach is to find a demixing or separating matrix $\mathbf{W}$ so that variables $\mathbf{u}_j$ in $\mathbf{U}$=$\mathbf{WX}$ are estimates of $\mathbf{S}_j$ up to scaling and permutation. In the deflationary approach, the sources are estimated one by one, by finding a column vector $\mathbf{w}_j$ (this will be stored as a row of $\mathbf{W}$) such that $\mathbf{u}_j = \mathbf{w}_j^T\mathbf{X}$ is an estimate of $\mathbf{s}_j$. Hence, $\mathbf{W}$ is an estimate of the (pseudo) inverse of $\mathbf{A}$ up to scaling and permutation of the rows of $\mathbf{W}$. In this work, the estimation of the independent components and the mixing matrix is done with the help of FastICA [95] in order to reduce the computational cost and the choice of the parameters of convergence. The FastICA is an iterative fixed-point algorithm that maximizes the non-gaussianity as a measure of independence, with the following update for $\mathbf{w}$:

$$\mathbf{w}(k) = \quad E\{\mathbf{X}g(\mathbf{w}(k-1)^T\mathbf{X})\} - \quad E\{g'(\mathbf{w}(k-1)^T\mathbf{X})\}\mathbf{w}(k-1) \tag{8.8}$$

where $\mathbf{w}$ is one of the rows of the demixing matrix $\mathbf{W}$, and g is the derivative of the contrast function, chosen to be a cubic polynomial. After each iteration step, $\mathbf{w}$ is normalized to have unit norm, ensuring that the rows $\mathbf{w}_j$ of the demixing matrix are orthogonal. The iteration is continued until the direction of $\mathbf{w}$ does not change significantly. With this procedure, the basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m\}$ is obtained and is used to represent each image as in equation 8.2.

8.1.2.3 Results and Discussion

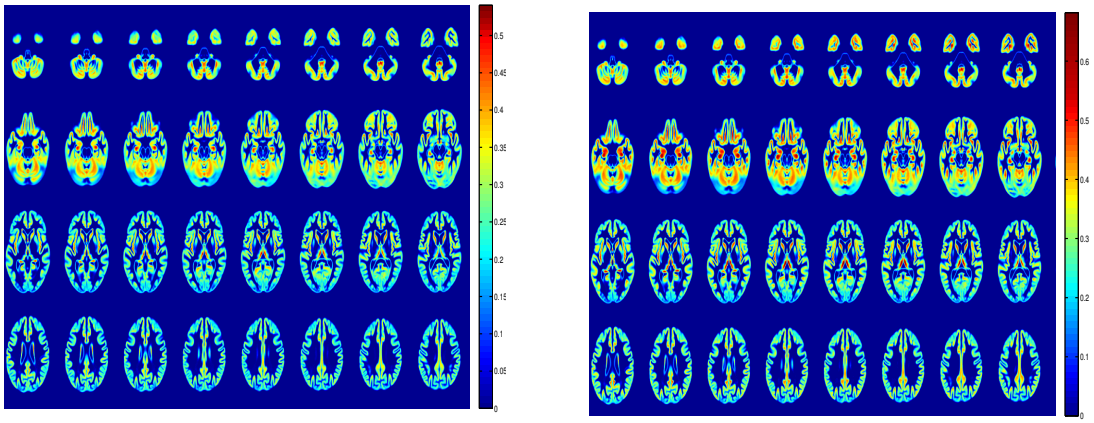
Three different experiments were performed considering the three different groups to construct the representative subspace $\mathbf{S}$ described in equation 8.4. In addition, the feature vector dimension (m) was determined by the number of the different average vectors classes.
Table 8.2 expands the classification results in the held out set for the case of m= 2

(considering NC and AD average vectors, NC and MCI average vectors or MCI and AD average vectors), once the described steps, projection, ICA has been applied to the raw data.
All the experiments represent a high compression of brain image data to a small number of features (m= 2 features). This proposed method represents a high compression of the large amount of information (voxels) contained in the brain images database to a small number of features. The results summarized in table 8.2 reveal that the idea of using SVM with non-linear kernel and a small number of features generalize better results than linear kernel. These obtained results confirm the theoretic reasons in [23, 129].

Table 8.2: Statistical performance measures of the proposed CAD system based on ICAm model.

Type of groups	Brain tissues	Kernel function	Accuracy	Sensitivity	Specificity	PL	NL
	GM	SVM-Lin	**0.881**	**0.909**	**0.851**	**6.101**	**0.106**
		SVM-RBF	**0.895**	**0.924**	**0.866**	**6.895**	**0.087**
NOR.vs.AD	WM	SVM-Lin	**0.786**	**0.786**	**0.786**	**3.673**	**0.276**
		SVM-RBF	**0.799**	**0.796**	**0.802**	**4.021**	**0.254**
	GM+WM	SVM-Lin	**0.872**	**0.909**	**0.845**	**5.864**	**0.107**
		SVM-RBF	**0.866**	**0.887**	**0.844**	**5.685**	**0.134**
	GM	SVM-Lin	**0.786**	**0.813**	**0.761**	**3.402**	**0.246**
		SVM-RBF	**0.796**	**0.829**	**0.765**	**3.527**	**0.223**
NOR.vs.MCI	WM	SVM-Lin	**0.739**	**0.781**	**0.696**	**2.569**	**0.315**
		SVM-RBF	**0.741**	**0.754**	**0.731**	**2.803**	**0.336**
	GM+WM	SVM-Lin	**0.776**	**0.813**	**0.739**	**3.115**	**0.253**
		SVM-RBF	**0.783**	**0.831**	**0.741**	**3.208**	**0.228**
	GM	SVM-Lin	**0.858**	**0.856**	**0.861**	**5.158**	**0.167**
		SVM-RBF	**0.855**	**0.882**	**0.831**	**5.218**	**0.142**
MCI.vs.AD	WM	SVM-Lin	**0.813**	**0.841**	**0.786**	**3.929**	**0.202**
		SVM-RBF	**0.831**	**0.866**	**0.796**	**4.245**	**0.168**
	GM+WM	SVM-Lin	**0.864**	**0.866**	**0.861**	**6.231**	**0.156**
		SVM-RBF	**0.858**	**0.876**	**0.839**	**5.441**	**0.147**

In addition, these results demonstrate that the idea of finding some representative images for the characterization AD is suitable.
Figure 8.4 represents the x_1 and x_2 values obtained when projecting each image into the AD and NC mean images (eq. 8.2).
Figure 8.5, 8.6 and 8.7 show an SVM classifier with linear kernel and the training vectors in the feature space (x_1, x_2), obtained by the projection of each image vector into the independent component $(\mathbf{u}_i)$ images space for the group 1, the group 2 and the group 3 respectively.
From the slope of the classifier for each group, the number of support vectors and the spread of the training data x_1 and x_2, a visual analysis can be performed to deduce the relevance of each pattern of atrophy for each image.



(a) AD Average of GM images

(b) NC Average of GM images

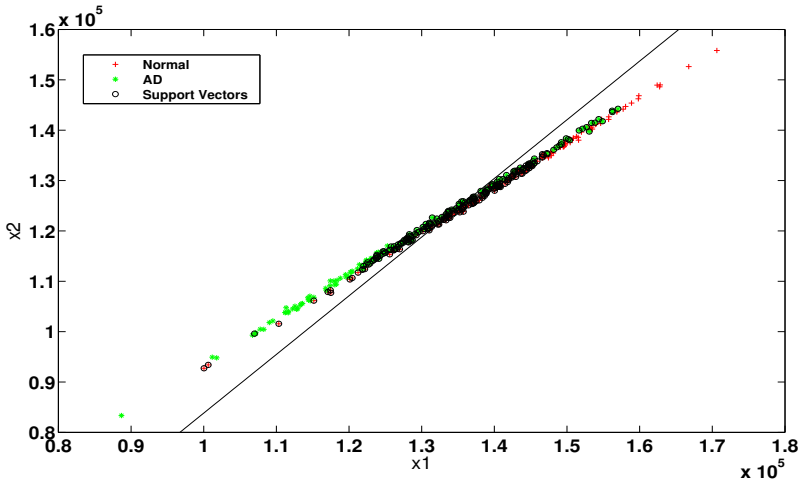
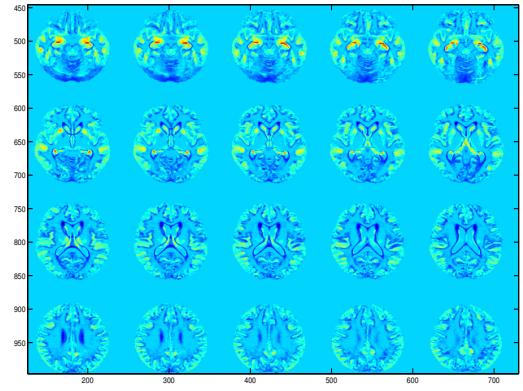
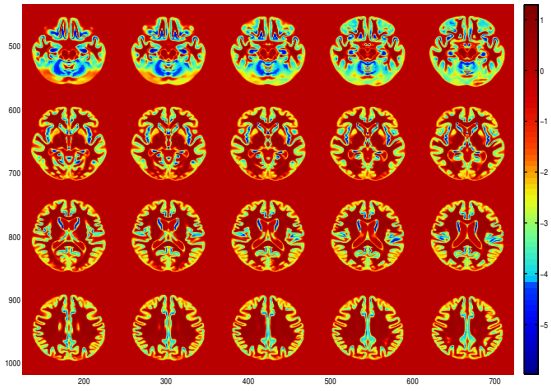


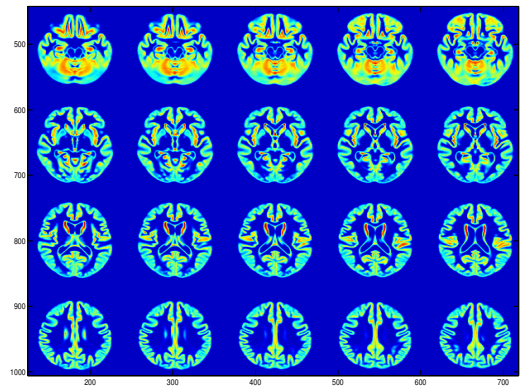
Figure 8.4: Projection of each GM image onto the NC and AD average image without the feature extraction method.



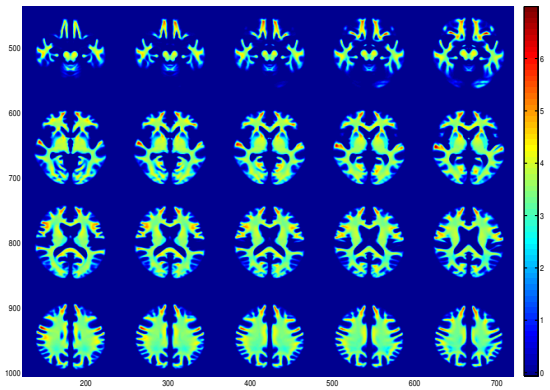
(a) $\mathbf{u}_1$ of GM images



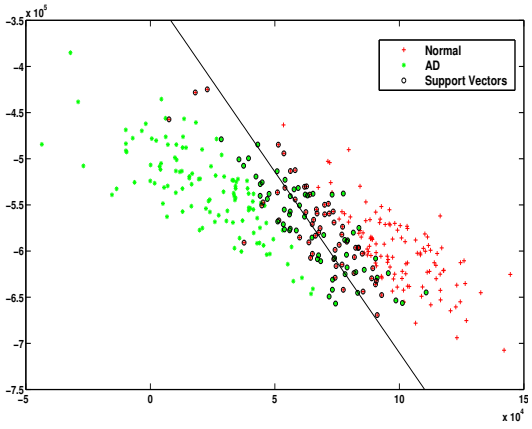
(b) $\mathbf{u}_1$ of WM images



(c) $\mathbf{u}_2$ of GM images



(d) $\mathbf{u}_2$ of WM images



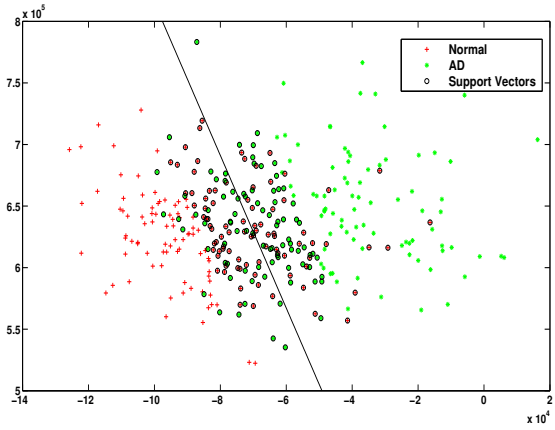
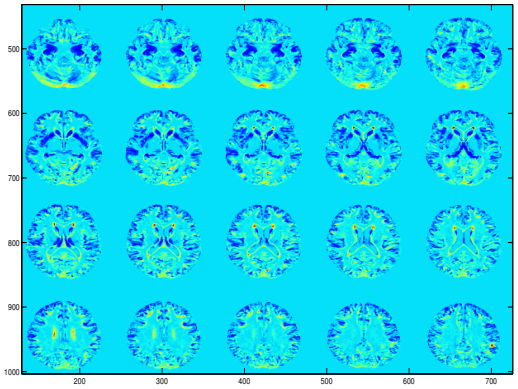
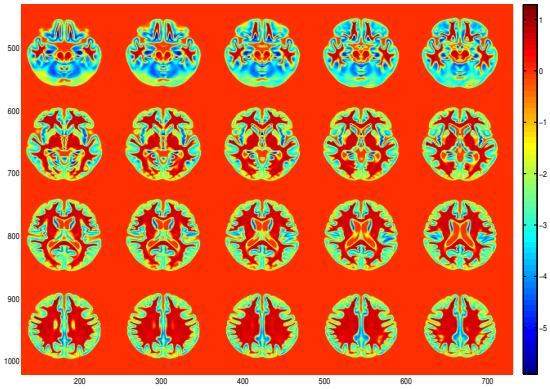


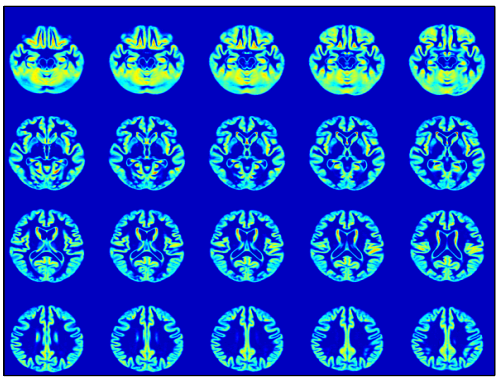
Figure 8.5: Separation of the two different classes for group 1 by projection of each GM images (left) and WM images (Right) into the image space, using linear SVM classifier.



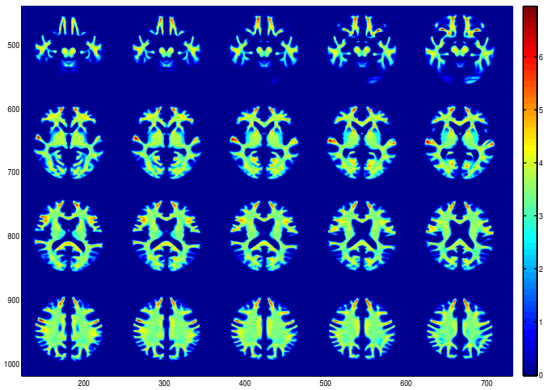
(a) $\mathbf{u}_1$ of GM images



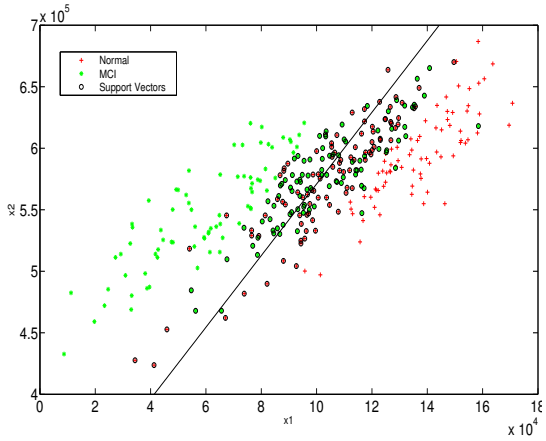
(b) $\mathbf{u}_1$ of WM images



(c) $\mathbf{u}_2$ of GM images



(d) $\mathbf{u}_2$ of WM images



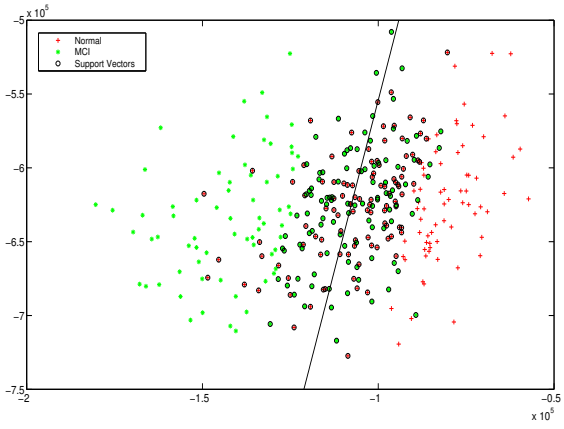
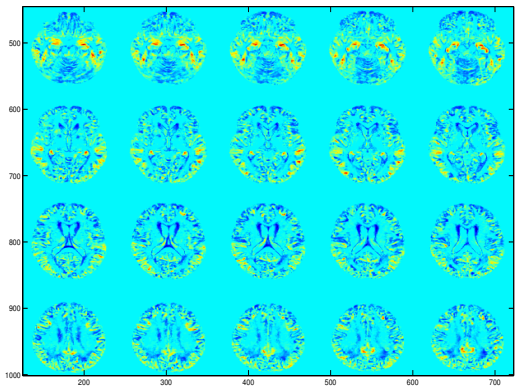
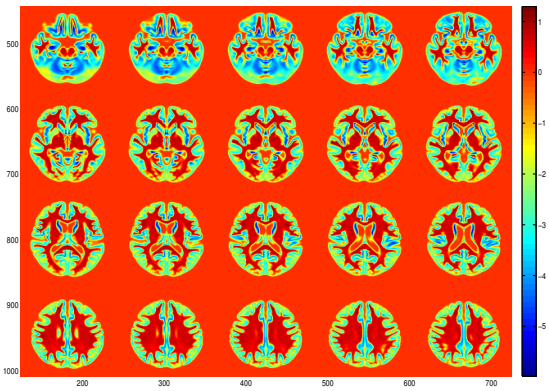


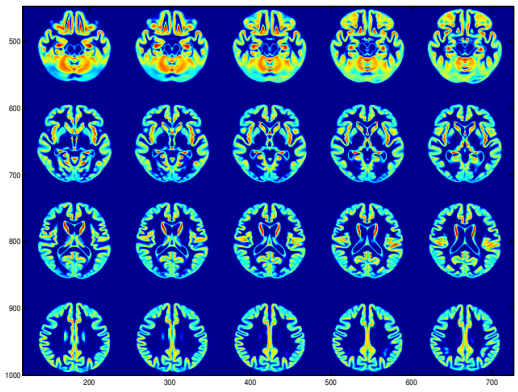
Figure 8.6: Separation of the two different classes for group 2 by projection of each GM images (left) and WM images (Right) into the image space, using linear SVM classifier.



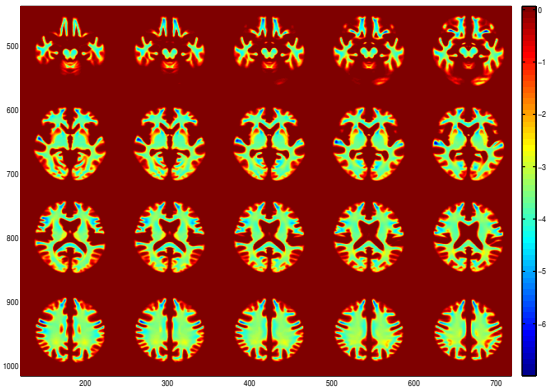
(a) $\mathbf{u}_1$ of GM images



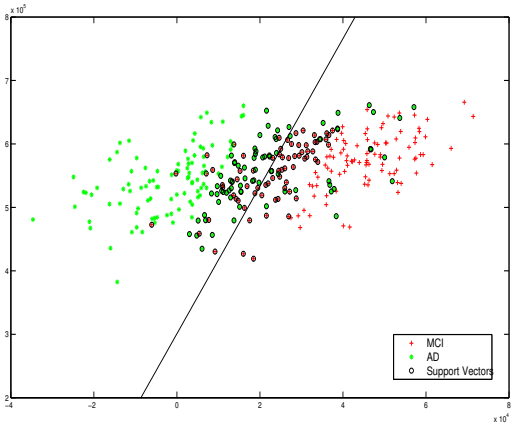
(b) $\mathbf{u}_1$ of WM images



(c) $\mathbf{u}_2$ of GM images



(d) $\mathbf{u}_2$ of WM images



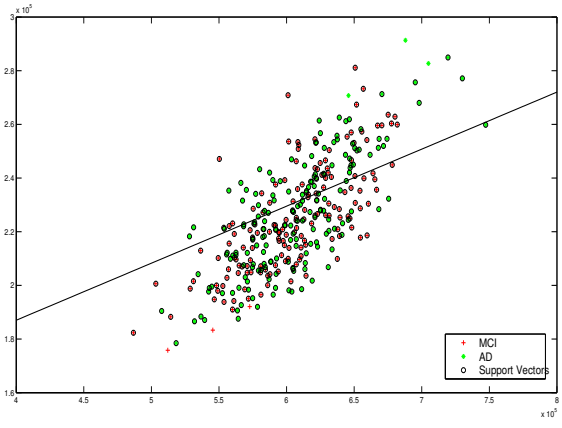


Figure 8.7: Separation of the two different classes for group 3 by projection of each GM images (left) and WM images (Right) into the image space, using linear SVM classifier.

Regarding the results of SVM classifier for the three groups with the different brain tissues, it can be noted that the separation of the GM tissue images in two classes is more efficient. This could be justified according to the existence of the most relevant information relies with the progress of the AD in the GM brain tissue. However, the SVM classification provides also good results with the WM images. It means that even in the early stage of the disease the WM images contain a discriminating features of AD. For this reason, it is better to consider both brain tissue distribution in analyzing the AD in the early stage.
The aims of this proposed methodology were, in the one hand, to compress the brain image dataset to a small element image basis for characterizing the AD in order to solve the problem of small sample size and to improve the computational time. On the other hand, to develop the coherent work flow for a CAD system for early AD with a meaningful decision support.
All the classification experiments were performed considering the three different groups for k-fold cross validation. Results for NOR/AD, NOR/MCI and MCI/AD classification are shown in table 8.2. In addition, group 1 shows more stable behavior and higher accuracy values than groups 2 and group 3. Thus, accuracy values up to 88% and sensitivity values up to 90% were obtained for group 1 with GM images. On the other hand, the classification results between MCI and AD provide 84% of accuracy and 85% of sensitivity levels. This obtained result reflects that the proposed CAD system has a good ability to distinguish between MCI subjects and patients with advanced AD. However, group 2 (NC.Vs. MCI) shows a lower accuracy values than groups 1 and 3. These low accuracy values may be explained by the use of imbalanced training data sets in group 2 (401 MCI and 229 NC). Therefore, the performance parameters can be affected by the introduction in groups 2 of MCI patients, whose image pattern of brain atrophy is complex and highly variable, and it evolves in time as the disease progresses [150, 163–165] (see figure 8.8). However, the introduction of MCI patients in group 3 (MCI. vs. AD) keep the stability of the statical measures.
Figure 8.9 depicts the ROC curves obtained by the proposed CAD system for the different brain tissues (GM, WM, GM+WM) of the sMRI database used in this thesis work. All the curves are very close to the upper left corner, so we have a desirable trade-off between sensitivity and (1 - specificity) in the area of interest. Furthermore, this figure confirms the higher accuracy of the proposed methodology and shows that it provides an adequate trade-off between sensitivity and specificity.

The analysis of the ROC curves shows that the proposed CAD system based on ICA model using GM images gives better result than the proposed approach using the combination of GM and WM images, mostly in the case of group 1. A possible interpretation of this difference may be that the atrophy in GM brain tissue is more acute, and the loss of GM is starting to be more visible for AD patients than in cases involving MCI subjects, where subtle differences may involve the whole GM and WM brain tissues [166].

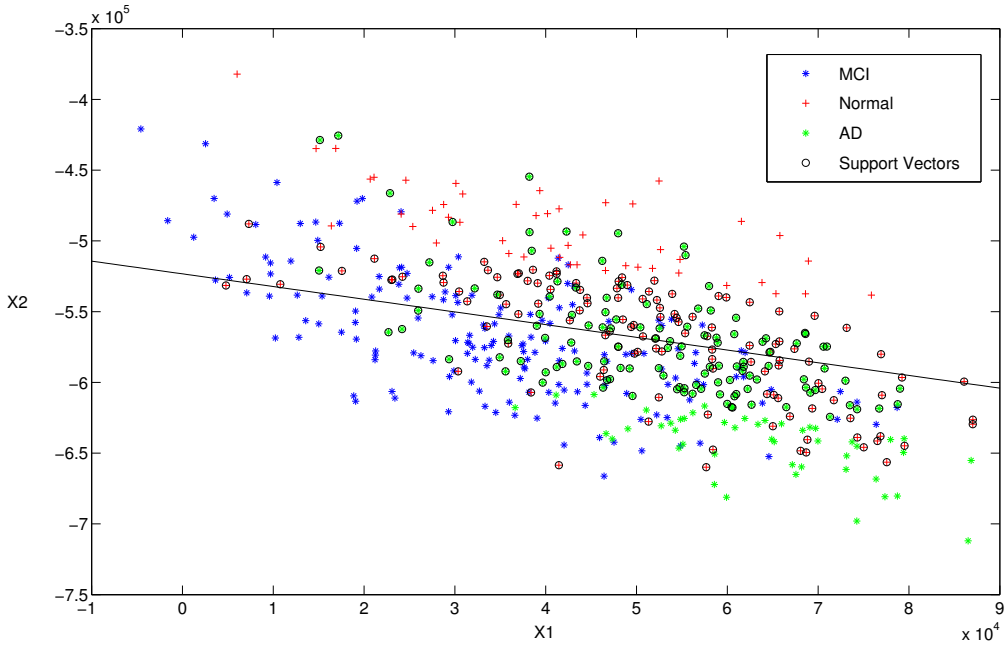
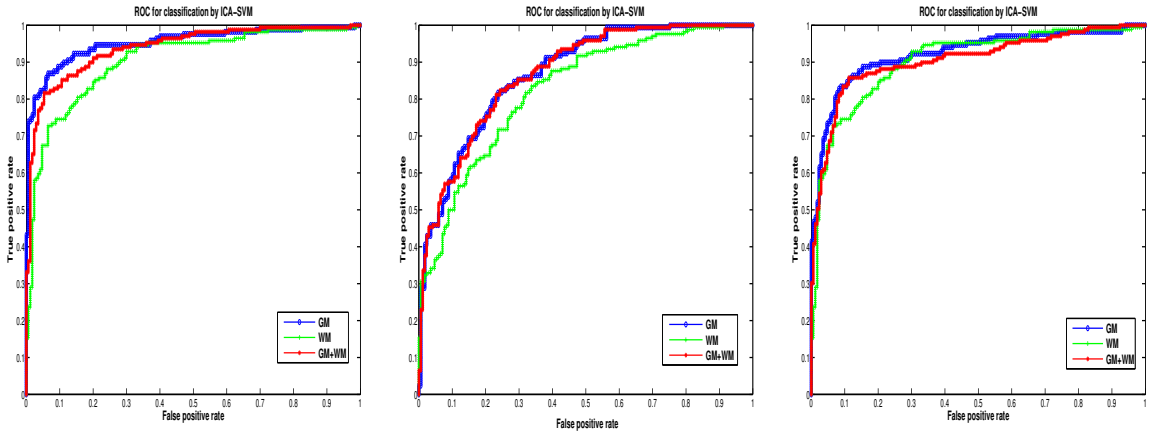


Figure 8.8: NC. vs. AD classification task: line decision designed by an SVM classifier with linear kernel when MCI subjects are included to design the classification rule.



(a) NOR. vs. AD (b) NOR. vs. MCI (c) MCI. vs. AD

Figure 8.9: Comparison of the trade off between sensitivity and specificity. ROC curves for the three cases analyzed: using only GM images, using only WM images and using both GM images and WM images.

- ***Visual support:***

The proposed CAD system is not only optimized in performance terms, but it also provides a meaningful interpretation of the classification output by means of eigenbasis projections onto 2D scatter plots. In figure 8.4, a two-dimensional basis formed by mean images is taken as candidate for the representative subspace. The projection onto the corresponding spanned subspace does not offer a visual support for classification. One needs to decompose each image by projecting it onto a more discriminative subspace.
In this way, there are several approaches in which this decomposition can be achieved. The most straightforward approach makes use of PCA [167]. But, as it occurs in the face recognition problem, the true dimensionality of the complete subspace [168] is typically not graphically representable. Moreover, in PCA, the basis of such subspace is not independent but uncorrelated.
To obtain an independent set of vectors or, at least, the most independent set of vectors, one should make use of ICA. This approach has been successfully applied to the problem of AD diagnosis with MRI [169], but the number of sources is again not representable, with the additional drawback of not having a natural sorting criteria and possible compensating scores.
Figures 8.5, 8.6, and 8.7 illustrate the representative patterns for each classification, together with the scatter-plot. The independent sources that serve as basis for equation 8.2 show a correspondence with previously reported patterns for the neurodegeneration process in AD ([170–173]).
Therefore, regions as the parahippocampal gyrus, lingual gyrus, hippocampus, frontal lobe, precentral gyrus, or the temporal lobe appear marked in the ICA-brain images. This information allows to decompose the test image into its class - representative patterns and to evaluate visually the amount of the atrophy of the image under study, by its location on the subspace projection. It also allows to detect any failure on the preprocessing steps, as normalization or segmentation, as the test image would appear as an outlier in the scatter-plot frame.
When average image of MCI are added to those of AD and NC, a 3-dimensional feature vector is produced (m=3). The three independent components are obtained in addition to $\mathbf{u}_1$ and $\mathbf{u}_2$ of figure 8.5.
Figure 8.8 demonstrates how the information of MCI is introduced in the classification task and also its corresponding coordinates for the training vector of group 1 (NC.vs. AD). The availability of a visual interpretation allows for a deeper understanding of the MCI case. For example, it contradicts the hypothesis of the work by Filipovych and Davatzikos [174], as it is shown that the MCI cases are not between NC and AD, but includes a more complex pattern of atrophy.
Generalization performance has been shown to be overestimated by cross-validation in many cases[175], and the appropriateness of an algorithm has been shown to rely on external factors to the algorithm itself, as the database size or acquisition[176], making controversial the direct comparisons between published results. In a recent 2016 work for the diagnosis of MRI[177], Harper and coworkers state that: "Visual assessment remains the primary method of scan interpretation, but in the absence of

a structured approach, diagnostically relevant information may be underutilized". In their work, a machine-learning approach is presented as an alternative to visual assessment, in order to compare both. The proposed approach in this work, that is an ICAm-SVM CAD, can be hardly evaluated under this premise, and uncertainly better than other CAD systems that make use of other features. The main advantage of the proposed CAD is that it produces a representable analysis of the classification of unseen data, by situating them in context, relative to other known cases. In contrast, other ICA-based CAD systems do not offer this structured approach to scan interpretation, as is the case of [169], or [178], where only group patterns can be visualized and unseen data are not related to them. Moreover, ICA used as feature extraction usually requires some feature selection criteria, in order to sort or select the independent components, as in [169, 178, 179] or [180], and are exposed to the randomness of ICA extraction. Feature selection may play a critical role in these works, making it hard to evaluate the ICA extraction effect. Here, we propose a solution to those problems by a modeled ICA extraction and projection, in the framework of eigenbrain decomposition [22]. This point of view also opens the possibility of future studies in the same sense of Harper et. al. in which visual support of CAD systems could be systematically evaluated.

8.1.3 Comparison between the both proposed methods

The results of the second proposed method are compared with the ICA model (method I), and also with the voxel-as-features (VAF) [181] baseline method, as showing in table 8.3.
The VAF method uses all voxels in each image as a feature vector, which is used as an input to the classifier. This approach is considered as a reference, because of its simplicity and the accurate estimation of its performance: in the work by Kloeppel et. al., [181] a pathological confirmed database was used to estimate a performance superior to visual assessments by experts.
In these experiments, only SVM linear has been used in the VAF approximation to compute the results, due to the large number of input features to the classifier, to obtain more generalizable results and to avoid the small sample size problem.
In light of the results shown in table 8.3, the second proposed approach improves the performance metrics of the analyzed CAD system, compared to VAF approach and the ICA approach (method I).

Table 8.3: Comparison betwwen the performance parameters achieved with ICA+LSVM, ICAm+LSVM, and VAF.

Type of groups	Brain tissues	Method	Accuracy	Sensitivity	Specificity	PL	NL
	GM	VAF	0.656	0.729	0.583	1.748	0.465
		Method I	0.846	0.865	0.824	4.915	0.164
		Method II	**0.881**	**0.909**	**0.851**	**6.101**	**0.106**
	WM	VAF	0.645	0.708	0.583	1.698	0.498
		Method I	0.703	0.729	0.671	2.216	0.404
NOR.vs.AD		Method II	**0.786**	**0.786**	**0.786**	**3.673**	**0.272**
	GM+WM	VAF	0.657	0.751	0.562	1.714	0.443
		Method I	0.864	0.884	0.839	5.491	0.138
		Method II	**0.872**	**0.909**	**0.845**	**5.864**	**0.107**
	GM	VAF	0.552	0.551	0.583	1.321	0.771
		Method I	0.695	0.691	0.699	2.295	0.442
		Method II	**0.786**	**0.813**	**0.761**	**3.402**	**0,246**
	WM	VAF	0.448	0.511	0.395	0.844	1.237
		Method I	0.635	0.633	0.638	1.748	0.575
NOR.vs.MCI		Method II	**0.739**	**0.781**	**0.696**	**2.569**	**0.315**
	GM+WM	VAF	0.521	0.562	0.479	1.078	0.914
		Method I	0.692	0.703	0.682	2.211	0.435
		Method II	**0.776**	**0.813**	**0.739**	**3.115**	**0.253**
	GM	VAF	0.489	0.458	0.521	0.956	1.041
		Method I	0.692	0.703	0.682	2.211	0.435
		Method II	**0.858**	**0.856**	**0.861**	**6.158**	**0.167**
	WM	VAF	0.595	0.649	0.525	1.366	0.668
		Method I	0.595	0.622	0.568	1.439	0.665
MCI.vs.AD		Method II	**0.813**	**0.841**	**0.786**	**3.929**	**0.202**
	GM+WM	VAF	0.611	0.667	0.534	1.431	0.623
		Method I	0.698	0.734	0.663	2.178	0,401
		Method II	**0.864**	**0.866**	**0.861**	**6.231**	**0.156**

In table 8.2, group 1 shows more stable behavior and higher accuracy values than groups 2 and 3. Thus, accuracy values up to 87% and sensitivity values up to 90% were obtained for group 1 with GM images, outperforming the classification results obtained when features are computed by means of VAF approach and the first proposed ICA model. This result is comparable to best of 10 methods in Cuingnet et. al.[121], in which their Voxel-Direct-D-GM reported best results in group 1, with sensitivity 85% - specificity 91% and was tested in the same database. Besides, the values of PL are more than the value considered to be a good class discriminant, and the NL are less than 0.2, considered also a good value.

8.2 Conclusion

For early AD diagnosis, new CAD systems based on two models of ICA were developed in this chapter. The first one is based on the t-test and ICA, as a feature selection and extraction techniques, respectively, to select the ROI and to extract a low number of IC which represent the feature vector for each image. The second model, called ICAm model, aims to extract the highly representative features from each average brain image (NC, MCI and AD), related to typical AD patterns, for classification and decision support. Thus, this model is implemented in a supervised way, such that an underlying model of the disease stages is used to extract the ICs as an eigenbrain decomposition, making feature selection unnecessary. The resulting CAD systems perform significantly well on the segmented sMRI images. In addition, these systems demonstrate its ability and robustness in AD detection and provide high accuracy values when using the first eight components. However, the ICAm model outperforms several proposed methods, such us, the baseline VAF approach and the first proposed ICA model.
As a conclusion, the developed ICA-based systems may be useful for computer aided diagnosis of AD. However, the generalization capabilities of the methodology could be overestimated and should be taken with caution. First, the ADNI is a multicenter database (approximate 50 centers using different voxel sizes and acquisition parameters). Second, scanner or center effects did not take into account in the experimental task. Third, potential brain vascular lesions in the subjects may be a confounding factor. Finally, the factors of age and gender have not been taken into account. All of these aspects may influence the final classification accuracy. [169].

Chapter 9

Structural MRI analysis based on Non-Negative Matrix Factorization

Non-Negative Matrix Factorization (NNMF) is a recently developed technique for finding reduced linear representations of non-negative data [113, 115, 182] being a useful decomposition tool for multivariate data.
NNMF decomposes the input data as a product of two matrices that are constrained by having nonnegative elements. It is firstly proposed by Lee and Seung [113] to decompose human face images and they achieved meaningful part-based representation due to only additive, not subtractive combinations are allowed. Recently, NNMF has been successfully applied to the biomedical applications and specifically in feature extraction and dimensionality reduction tasks.
The mathematical details of this approach are presented in chapter 4. Furthermore, all the obtained results during this work with the different proposed experiments are presented in the following.

9.1 Experiments

The normalized and segmented sMRI database is selected as input data for the NNMF + SVM CAD tool. Each patient has 125 × 145 × 125 voxels, which yields 2265625 voxels per patient. Then, a binary mask is applicable by computing the average of all the normal subject brain images. Only the voxels that have an intensity above 10% of the maximum intensity in the average image will be considered. This step reduces the number of voxels in the input space. After that, the NNMF approach is applied in order to select the most relevant features related to AD. Finally, the SVM classifier is applied to distinguish between NC, MCI and AD. The pipeline of this developed CAD system is illustrated in the figure 9.1.

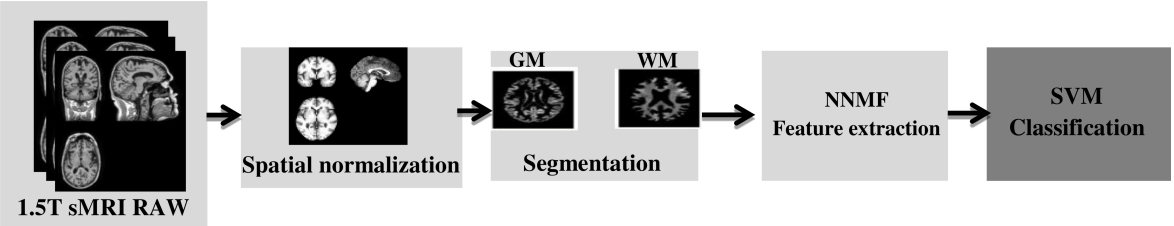


Figure 9.1: Illustration of the NNMF-CAD system.

9.2 Results

This section provides the experimental results of the evaluation of the developed CAD system, along with its variants. The NNMF+SVM based CAD tool is developed with linear and non-linear SVM classifier. Thus, the classification results are summarized in table 9.1 with different experiments (GM images, WM images and the combination of feature extracted from GM and WM segmentation).
Figures 9.2, 9.3 and 9.4 show the NMF + SVM results in terms of accuracy (Acc), sensitivity (Sens) and specificity (Spec), for the different tissue brain (GM, WM and GM+WM) with the different SVM kernels. According to these figures, the best classification results are with linear SVM kernel. This proposed methodology using NNMF as feature extraction technique yields a peak values of accuracy = 89.04%, sensitivity = 88.23% and specificity = 89.83%.

Table 9.1: Statistical measures of performances of NNMF+SVM approach with the different kernels.

	Linear				RBF		
Groups	Tissue brain	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
NOR. vs. AD	GM	85.1064%	87.766%	82.4468%	85.9043%	88.8298%	82.9787%
	WM	82.9787%	81.9149%	84.0426%	81.383%	83.5106%	97.2553%
	(GM+WM)	**89.0391%**	**88.234%**	**89.8298%**	**86.1702%**	**83.5106%**	**88.8298%**
NOR. vs. MCI	GM	73.6702%	75%	72.34%	73.1383%	76.0638%	70.2128%
	WM	76.3298%	79.7872%	72.8723%	71.5426%	72.8723%	70.2128%
	(GM+WM)	**80.0532%**	**81.9149%**	**78.1915%**	**74.4681%**	**76.0638%**	**72.8723%**
MCI. vs. AD	GM	71.8085%	71.2766%	72.3404%	66.2234%	64.3617%	68.0851%
	WM	75.7979%	76.0638%	75.5319%	75.266%	76.0638%	74.4681%
	(GM+WM)	**79.5213%**	**79.2553%**	**79.7872%**	**73.1383%**	**70.2128%**	**76.0638%**

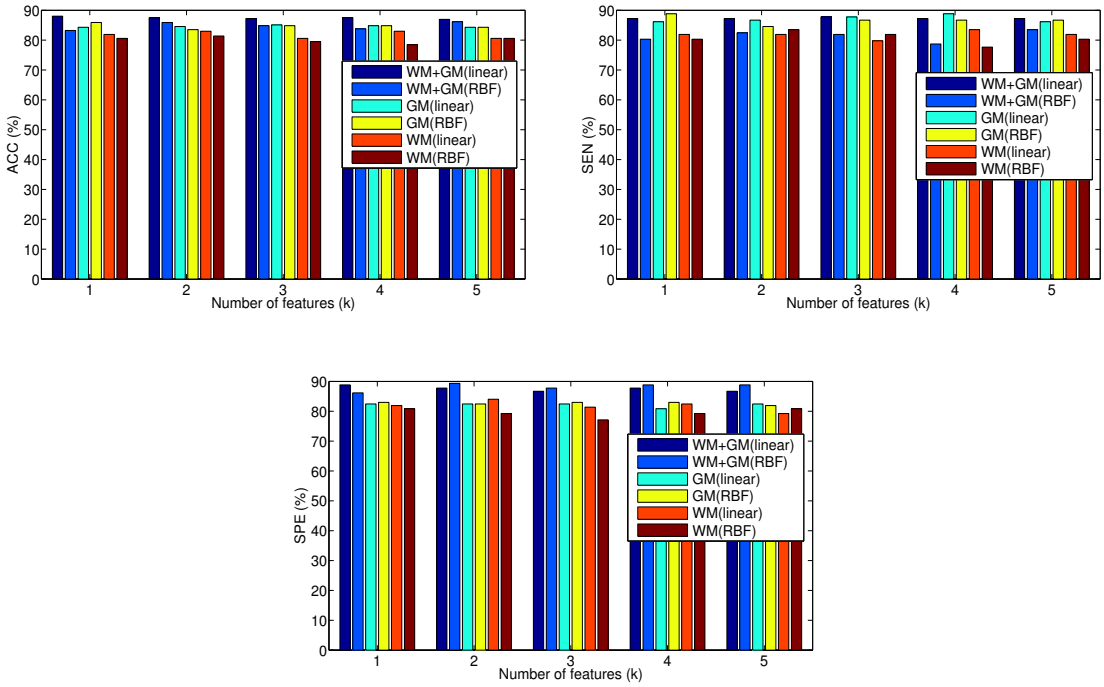


Figure 9.2: Performance of the NNMF-SVM CAD system with the segmented MRI database, for different k values, only NOR/AD groups are considered.

9.3 Discussion

This work shows a CAD system for the early detection of the Alzheimer's disease. A NNMF + SVM analysis for segmented structural brain images is applied for the proper classification of MRI images and AD detection. The proposed methodology is based on NNMF for selection and extraction of the MRI image features of each patient of the database, considered as source data for further classification through an SVM-based method, considering different kernel functions.
In table 9.2, the developed CAD system that is presented in this chapter was compared with other existing in the literature. The voxel-as-features (VAF) approach results are reported as reference. In all the groups, the proposed methodology outperforms the VAF approximation method [19]. As a conclusion, our methodology can produce a valid approach to perform a CAD system for early diagnosis of AD.

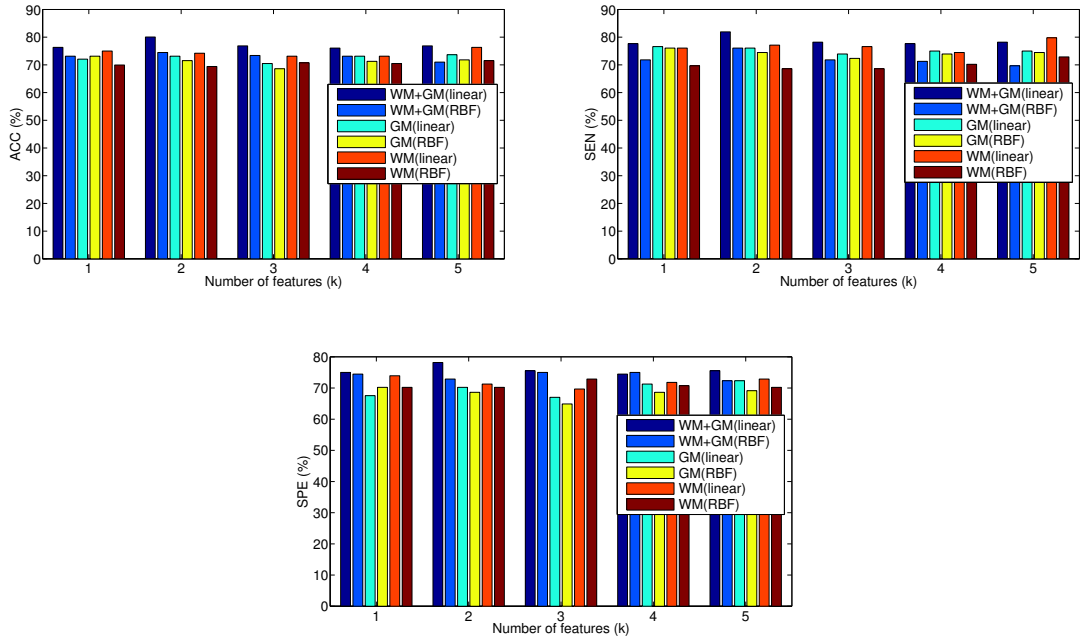


Figure 9.3: Performance of the NNMF-SVM CAD system with the segmented MRI database, for different k values, only NOR/MCI groups are considered.

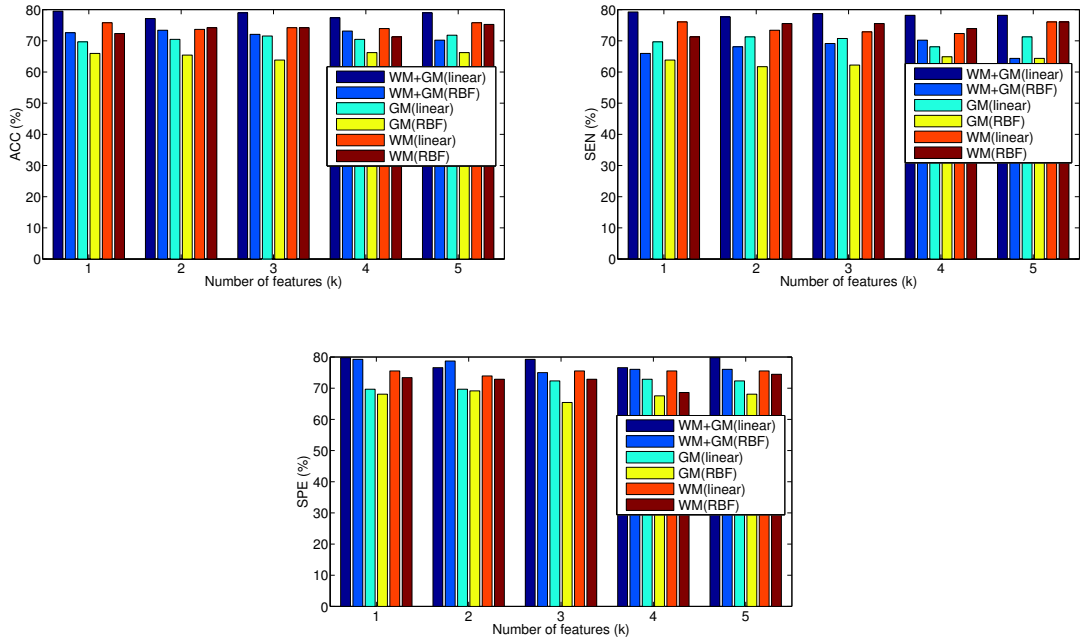


Figure 9.4: Performance of the NNMF-SVM CAD system with the segmented MRI database, for different k values, only MCI/AD groups are considered.

Table 9.2: Comparison of performance parameters using VAF and the proposed NNMF model by means of SVM with linear kernel.

Type of groups	Brain tissues	Method	Accuracy	Sensitivity	Specificity
	GM	VAF	0.656	0.729	0.583
		the proposed methodology	**0.881**	**0.909**	**0.851**
	WM	VAF	0.645	0.708	0.583
		the proposed methodology	**0.786**	**0.786**	**0.786**
	GM+WM	VAF	0.657	0.751	0.562
NOR.vs.AD		the proposed methodology	**0.872**	**0.909**	**0.845**
	GM	VAF	0.552	0.551	0.583
		the proposed methodology	**0.786**	**0.813**	**0.761**
	WM	VAF	0.448	0.511	0.395
		the proposed methodology	**0.739**	**0.781**	**0.696**
	GM+WM	VAF	0.521	0.562	0.479
NOR.vs.MCI		the proposed methodology	**0.776**	**0.813**	**0.739**
	GM	VAF	0.489	0.458	0.521
		the proposed methodology	**0.858**	**0.856**	**0.861**
	WM	VAF	0.595	0.649	0.525
		the proposed methodology	**0.813**	**0.841**	**0.786**
	GM+WM	VAF	0.611	0.667	0.534
MCI.vs.AD		the proposed methodology	**0.864**	**0.866**	**0.861**

The second experiment evaluates the classifier performance in terms of the Receiver Operating Characteristic (ROC) curves, that evaluate the trade-off between the false positive rate and the true positive rate of the ensemble.
Figure 9.5 reveals the ROC curves for increasing complexity classification tasks: NC. vs. AD, NC. vs. MCI and MCI. vs. AD, considering NNMF technique.

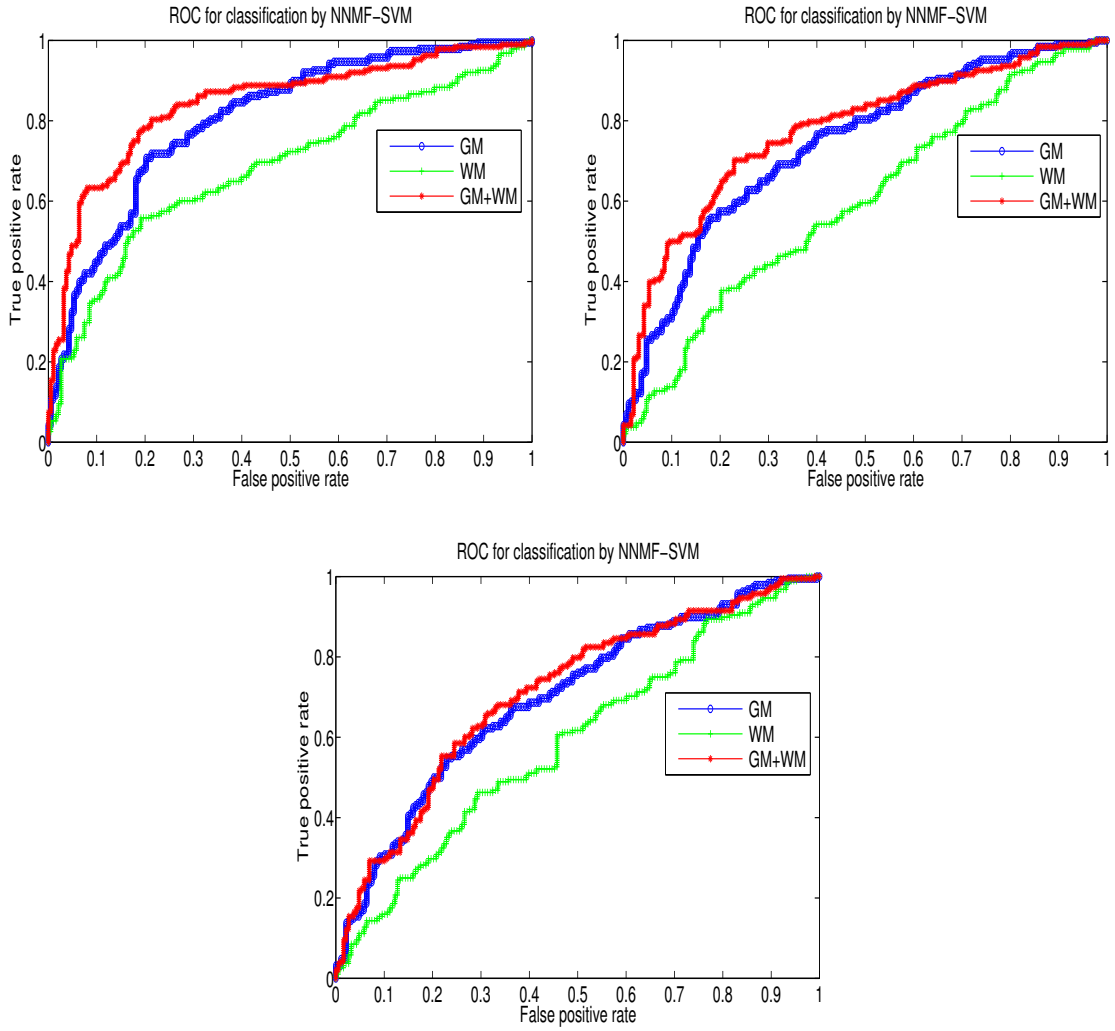


Figure 9.5: Comparison of the trade off between sensitivity and specificity. ROC curves for the three cases analyzed: using only GM images, using only WM images and using both GM images and WM images.

9.4 Conclusion

In this chapter, a novel image-based computer system is developed for an automatic detection and diagnosis of the early stages of AD. The proposed system is based on the combination of NNMF, as a feature reduction technique and the SVM classifier. The feature reduction step provides a reduced set of variables representing the original data. This dimensionality reduction approach is suitable, specifically, for machine learning techniques, such as, the SVM.
The NNMF-SVM CAD tool is validated on sMRI image database which provides information about the structural change in the brain. The validation results of the proposed NMF-SVM CAD system yields up to 89% classification accuracy with high sensitivity and specificity values (upper than 88%) when we distinguish between controls and AD classes. In performance terms, the developed system outperforms several proposed methods from the bibliography, specifically the baseline VAF approach.

Part III

General discussion and conclusions

Chapter 10

General discussion, conclusions and future work

The different contributions of this thesis have already been discussed in detail in each chapter. However, in the first section of this chapter, we will discuss and compare the diffe-rent image-based CAD systems, proposed for an accurate identification of Alzhei-mer's disease, specifically, in its early stage. Moreover, the majority of the developed CAD systems have been published in [180, 183–185].
The second section shows the conclusions of this PhD thesis and highlights the scientific contributions that have been made. These contributions are based on the development of new algorithms to improve the accuracy of computer aided diagnosis systems for neurodegenerative diseases, such as, Alzheimer's disease in its early stage. Furthermore, these CAD systems aimed to solve the sample size problem in neuroimaging. These proposed CAD systems achieve a satisfactory classification performance, being capable of distinguishing between normal, MCI and AD patients successfully.
A series of proposals are made to continue this work and to open new lines of research that will improve these methods, in the last section.

10.1 Comparison and discussion between the developed CAD systems

Different decomposition techniques for tackling the small sample size problem were developed in this PhD thesis. These techniques (chapters 6, 7, 8 and 9) are based on feature selection and feature extraction algorithms that perform a significant reduction of the number of features used in neuroimaging. The classification task is performed using an SVM classifier with linear and RBF kernels, and performance values are obtained via k-fold cross-validation.
Through the developed CAD systems, we obtained a very significant feature reduction, from hundreds of thousands of voxels to between 2 and 10 features, which were the coordinates of each sample in the space defined by the components. All the feature selection and feature extraction techniques which are used in this work were able to detect the region of interest in the AD structural datasets, obtaining an accuracy around 90%. These techniques make also our CAD systems better generalizable, since the features are no longer subject to the small sample size problem. Thus, the samples are projected to a dense space, where the SVM are able to perform a reliable classification.

Results of the different developed CAD systems are shown in table 10.1, 10.2 and 10.3 for classifying NOR/AD, NOR/MCI and MCI/AD groups. These results were compared to VAF technique, as a reference. Firstly, the segmentation process is performed on sMRI in order to analyze the features extracted from GM and WM images, separately. Afterwards, a combination of these features are realized to improve the accuracy rates of these developed systems, specifically, in the early stage of AD. Through the obtained results of the different proposed techniques, we have shown that combining features extracted from GM and WM segmentation gives a good classification accuracy than those using only GM brain tissue or WM brain tissue separately. Thus, this idea adds a valuable robustness to our developed systems for distinguishing AD and MCI from NC subjects. Secondly, all the developed systems were achieved the best accuracy results in the AD. vs. NC classification ($\simeq$ 89%). When the intermediate state MCI is involved, the performances are affected because of the complexity and the variability of brain atrophy patterns based on MRI. PLS + LSVM achieved the best accurate rates for NC. vs. MCI (81.89% of accuracy, 82.16% of sensitivity and 81.62% of specificity), and ICAm + LSVM achieved the best accurate rates for MCI. vs. AD (86.4% of accuracy, 86.6% of sensitivity and 86.1% of specificity). In light of these obtained results, the proposed techniques improve the performance metrics of the CAD systems, specially in the early stage of the AD, compared to VAF approach.

As a conclusion, we have proved the ability of all these developed CAD systems in the differential diagnosis of AD. All these proposed methodologies perform ei-

Table 10.1: Statistical measures of performance of the developed CAD systems with linear SVM classifier, for NOR/AD group, and using *k*-fold cross validation (k=10).

Type of groups	Brain tissues	Method	Accuracy	Sensitivity	Specificity
NOR.vs.AD	GM	VAF	0.656	0.729	0.583
		PLS-SVM	87.53	88.65	86.17
		PCA-SVM	85.61	89.08	81.38
		ICA-SVM	84.6	86.5	82.4
		ICAm-SVM	88.1	90.9	85.1
		NNMF-SVM	85.10	87.76	82.44
	WM	VAF	0.645	0.708	0.583
		PLS-SVM	85.61	87.34	83.51
		PCA-SVM	81.77	84.28	78.72
		ICA-SVM	70.3	72.9	67.1
		ICAm-SVM	78.6	78.6	78.6
		NNMF-SVM	82.97	81.91	84.04
	GM+WM	VAF	0.657	0.751	0.562
		PLS-SVM	88.49	91.27	85.11
		PCA-SVM	87.77	89.96	85.11
		ICA-SVM	86.4	88.4	83.9
		ICAm-SVM	87.2	90.9	84.5
		NNMF-SVM	89.03	88.23	89.82

ther a reduction of the feature space in order to reduce the amount of false positives currently found in neuroimaging studies. Other improvements, such as the computational load reduction thanks to feature extraction techniques may also be acknowledged.

10.2 Conclusions

In the first part of this work, different existing approaches for the diagnosis are reviewed. All the tools that will be used for further analysis and development of this PhD research work are introduced. The characteristics of AD and also the different clinical diagnostic methodologies for analyzing the structural tomographic images, are described. A specific chapter is devoted to the detailed description of the data used for validation of the developed diagnostic techniques. The fundamental concepts of the theory of statistical learning are introduced with special interest to the

Table 10.2: Statistical measures of performance of the developed CAD systems with linear SVM classifier, for NOR/MCI group, and using k-fold cross validation (k=10).

Type of groups	Brain tissues	Method	Accuracy	Sensitivity	Specificity
NOR.vs.MCI	GM	VAF	0.552	0.551	0.583
		PLS-SVM	77.57	76.76	71.9
		PCA-SVM	75.41	78.38	72.43
		ICA-SVM	69.5	69.1	69.9
		ICAm-SVM	78.6	81.3	76.1
		NNMF-SVM	73.67	75	72.34
	WM	VAF	0.448	0.511	0.395
		PLS-SVM	80.54	79.46	81.62
		PCA-SVM	75.14	77.3	72.98
		ICA-SVM	63.5	63.3	63.8
		ICAm-SVM	73.9	78.1	69.6
		NNMF-SVM	76.32	79.78	72.87
	GM+WM	VAF	0.521	0.562	0.479
		PLS-SVM	81.89	82.16	81.62
		PCA-SVM	78.92	80	77.84
		ICA-SVM	69.2	70.3	68.2
		ICAm-SVM	77.6	81.3	73.9
		NNMF-SVM	80.05	81.91	78.19

supervised learning methods applied to the small sample size problem compared to the dimension of feature space. Furthermore, several analysis tools, such as, the ROC curves and the cross validation method are implemented at the classification stage to evaluate the performance of the developed CAD systems.
In this sense, this PhD work provides new contributions based on the development and the implementation of new approaches for the analysis of segmented structural magnetic imaging (sMRI) brain images. More specifically, four approaches are developed, which can improve the accuracy of the CAD systems used for the diagnosis of AD.

The second part of this research work describes the main contributions through the development of four novel CAD systems for the early diagnosis of AD. These CAD systems are based on feature selection and feature extraction techniques to extract the significant voxel's intensity in the regions of interest (ROI). ROI was estimated using an image model called "template". This image model was computed as an average of control subject images, which leads to define the ROI as those brain areas

Table 10.3: Statistical measures of performance of the developed CAD systems with linear SVM classifier, for MCI/AD group, and using k-fold cross validation (k=10).

Type of groups	Brain tissues	Method	Accuracy	Sensitivity	Specificity
MCI.vs.AD	GM	VAF	0.489	0.458	0.521
		PLS-SVM	77.03	74.59	79.46
		PCA-SVM	72.71	72.43	72.97
		ICA-SVM	0.692	0.703	0.682
		ICAm-SVM	0.858	0.856	0.861
		NNMF-SVM	71.8	71.27	72.34
	WM	VAF	0.595	0.649	0.525
		PLS-SVM	87.03	88.65	85.41
		PCA-SVM	79.19	82.16	76.22
		ICA-SVM	0.595	0.622	0.568
		ICAm-SVM	0.813	0.841	0.786
		NNMF-SVM	75.79	76.06	75.53
	GM+WM	VAF	0.611	0.667	0.534
		PLS-SVM	85.41	87.03	83.78
		PCA-SVM	81.89	84.86	87.92
		ICA-SVM	0.698	0.734	0.663
		ICAm-SVM	0.864	0.866	0.861
		NNMF-SVM	79.52	79.25	79.78

that are most active in healthy people.

In Chapter 6, the first proposed methodology is presented, which makes use of the t-test and partial least squares (PLS) approaches. The first approach is used to select the voxels, where the structural tissues differ from healthy and patient subjects. The second approach decomposes the data in two sets of variables into the product of two matrices, called scores and loadings. In our case, the two sets of variables are those formed by the segmented structural images and by labels of these images. After the decomposition of these sets, the score matrix corresponding to the set of images contains the projection of voxels which were obtained considering the labels of the images. These projections are known as PLS-brains which are used as feature vectors. These later used as input parameters for the classifier.
In chapter 7, a linear decomposition of the segmented data in eigenvectors called principal components (PCs) through the principal component analysis (PCA) technique were obtained. The combination of PCA, the support vector machines (SVM) and the t-test selection criterion have proved to be a suitable technique for neu-

rological image classification. Thus, it achieved a good classification result and it solved the small sample size problem.
Chapter 8 presents the obtained results by the two proposed CAD systems based on independent component analysis (ICA). The first CAD system uses a combination of voxel selection and extraction techniques based on t-test and ICA respectively. This combination allows us to extract the highly representative features, which are closely related to typical AD patterns. The second CAD system is based also on the ICA approach to extract the highly representative features, but from each average brain image for the different classes (NC, MCI and AD), related to typical AD patterns. However, this approach is implemented in a supervised way, such that an underlying model of the disease stages is used to extract the ICs as an eigenbrain decomposition, making feature selection process unnecessary. The last chapter presents all the results of the last developed CAD system using non-negative matrix factorization (NNMF) approach. This approach reduces the original data through a linear representation of only the non-negative data of variables.
The analysis of the segmented structural MRI images using these proposed approaches, aims to extract reduced feature vectors from gray matter brain images and white matter brain images, separately, to discriminate between AD, MCI and elderly normal control (NC) subjects. These feature vectors were used separately or together to train a statistical classifier. These approaches show a satisfactory overall performance with a classification accuracy of 89% when combining the features extracted from gray matter and white matter tissue brain. In this way, these approaches allow us to improve the precision of the CAD systems for diagnosing the AD in its different stages.
In the classification stage of proposed CAD systems, the SVM is used as a classifier with linear and non linear kernel (RBF) due to its high robustness to the final decision process. Therefore, the validation of these CAD systems has been estimated using the k-fold cross validation technique. The obtained results are compared with previous approaches like voxel-as-feature (VAF) approach. These results are detailed in Chapter 6, 7, 8 and 9. The main goals of the PhD project were successfully reached. This is confirmed by the fact that the works developed were published in several international journals and presented at international conferences.

10.3 Future work

As future lines of the research developed throughout this work, we present the following proposals:

- The use of Multiclass classification (NC, MCI and AD) that allows us to categorize the images in more than two classes (NOR/AD, NOR/MCI, MCI/AD) and therefore to distinguish the different stages of the Alzheimer's disease in

the same time.

- To investigate the case of discriminating "non converter-MCI" from "converter-MCI" patients, since this is one of the most clinically interesting challenges.

- Multiple biomarkers may provide complementary information for the diagnosis of AD. It would be interesting to see if incorporating the various modalities (clinical and neuropsycological assessment scores, longitudinal imaging, biological data) enables to improve between "non converter-MCI" and "converter-MCI" patients.

Part IV

Summary in Spanish

Chapter 1

Introducción

Este capítulo describe en primer lugar los objetivos y la organización de este resumen. En segundo lugar, una introducción a la enfermedad neurológica que nos ocupa en esta trabajo (Enfermedad de Alzheimer). Posteriormente, se explican la obtención y el preprocesamiento de la base de datos utilizada en este trabajo. En particular, utilizamos una base de datos sMRI de ADNI. Finalmente, la última sección de este capítulo presenta un resumen de los resultados experimentales obtenidos en esta tesis doctoral.

11.1 Objectivos y organización

Los objetivos principales de este trabajo se dirigen al desarrollo de técnicas novedosas de procesado y clasificación de imágenes estructurales cerebrales de resonancia magnética para la construcción de sistemas de ayuda a la toma de decisiones en el contexto clínico del diagnóstico de Enfermedades neurodegenerativas con especial atención a la Enfermedad de Alzheimer.
Los objetivos concretos del trabajo se resumen en el desarrollo de técnicas avanzadas de extracción de información y clasificación automática de imágenes estructurales de resonancia magnética (sIRM) en el contexto de identificación de pacientes afectados por la enfermedad de Alzheimer.
El diseño y la evaluación de los sistemas de diagnóstico asistido por ordenador (DAO) desarrollados empleará las bases de datos ADNI (Alzheimer's disease Neuroimaging Initiative).
Un estudio multicéntrico cuyo objetivo es evaluar los biomarcadores que describan el progreso de la enfermedade de Alzheimer. Por lo tanto, se proporciona las sMRI correspondientes a sujetos normales, pacientes con deterioro cognitivo (MCI) o de enfermedad de Alzheimer para la detección de la enfermedad en su fase precoz.

El informe de doctorado empezó para la descripción de la enfermedad de Alzheimer, los efectos en nuestra sociedad y la importancia del diagnóstico precoz.

El capituló 3 describe las bases de datos sobre las que vamos a trabajar, así como la etapa de pre-procesamiento que se utiliza para mejorar la calidad de la imagen y también para mejorar los resultados de detección de la enfermedad en su fase precoz.

En el capituló 6, 6, 6 y 6 se describen cuatro nuevas técnicas para desarrollar el diagnóstico, basado en los mínimos cuadrados parciales (PLS), componentes principales (PCA), técnicas de componentes independientes (ICA) y la matriz de factorización no negativa (NNMF) enfoques. Las conclusiones y las perspectivas de futuro se explican en el capítulo 10.

11.2 La Enfermedad de Alzheimer

La Enfermedad de Alzheimer (EA) es una de las enfermedades neurodegenerativas más comunes entre las personas mayores y produce dramáticas consecuencias socioeconómicas y para la salud [1]. Con el envejecimiento de la población en los países desarrollados, la relevancia de esta enfermedad aumenta cada día.

Los tratamientos disponibles actualmente minimizan el impacto de la enfermedad

y se necesitan intervenciones más efectivas. Lograr esto requerirá una comprensión sofisticada de su causa o causas. La definición de EA ha variado a lo largo de los años. Inicialmente y varias décadas después, EA fue considerado una forma de demencia presenil. Los individuos dementes mayores no fueron diagnosticados generalmente con EA, aunque sus cerebros con frecuencia contenían placas neuríticas y enredos neurofibrilares. Posteriormente, la definición fue cambiada para incluir a todos los pacientes que se consideraba que tenían una demencia relacionada con las placas y los enredos, independientemente de la edad [186]. La EA es una enfermedad crónica y progresiva que dura muchos años y, desafortunadamente, los datos controlados con placebo sobre los fármacos anti-EA son principalmente de estudios a corto plazo [187]. Los tratamientos de la EA podrían dividirse en tratamientos sintomáticos que mejoran los síntomas sin afectar el proceso de la enfermedad subyacente, y las terapias modificadoras de la enfermedad que afectan al transcuro de la enfermedad ralentizando la evolución de la misma [188].

11.3 las imágenes por Resonancia Magnética

La Imagen por Resonancia Magnética (IRM) es un proceso de formación de imágenes médicas, basado en el fenómeno conocido como resonancia magnética nuclear (NMR, del inglés Nuclear Magnetic Resonance), que está ganando aceptación generalizada para una gran variedad de exploraciones médicas [28]. Es una de las técnicas más utilizadas en radiología para obtener información sobre la estructura y composición del cuerpo a analizar, incluyendo el cerebro. Algunas de las razones para este creciente interés son:

- No es invasivo, basado en la radiación no ionizante, procedimiento de imagen.
- La rápida evolución de las técnicas de imagen de resonancia magnética que ofrecen un amplio repertorio de secuencias de pulsos que pueden ser fácilmente configuradas para ofrecer visualizaciones específicas. La IRM convertido en una herramienta de imagen muy flexible.
- Tiene una alta resolución espacial y proporciona mucha información sobre la estructura anatómica, lo que permite la cuantificación de la patológica y de los estudios clínicos.

Recientemente, una variedad de modalidades de imagen de resonancia magnética, incluyendo IRM estructural y funcional, han demostrado cambios característicos en el cerebro de pacientes con EA, en los estados prodrómicos e incluso los estados presintomáticos que pueden ayudar en el proceso fisiopatológico de EA. Sin embargo,

este trabajo se centra únicamente en la IRM estructural. Este último tiene una alta resolución espacial y proporciona mucha información sobre la estructura anatómica del celebro.

11.4 Los sistemas de diagnóstico asistido por ordenador

En el inicio tardío de la EA, hay un déficit de perfusión mínimo en las etapas de la enfermedad, y los cambios relacionados con la edad, que se ven con frecuencia en las personas de edad sanos, tienen que ser discriminados de los cambios mínimos específicos de la enfermedad. La detección de estos cambios mínimos en las imágenes de diagnóstico visual es una tarea difícil que necesita clínicos experimentados. Con el fin de mejorar la precisión de la predicción, especialmente en las primeras etapas de la enfermedad cuando el paciente podría beneficiarse más de los medicamentos y tratamientos, los sistemas de diagnóstico asistido por ordenador (DAO) son deseables. Varias tecnicas para el diseño de sistemas DAO de la EA se pueden encontrar en la literatura [21].
La metodología univariada se basa en el análisis de las regiones de interés (ROI) mediante algunas funciones discriminantes, mientras que el enfoque multivariable está relacionado con las técnicas de análisis estadístico. Además, el primer enfoque se basa en la herramienta de software Statistical Parametric Mapping (SPM) [18]. Este enfoque no fue desarrollado específicamente para estudiar una sola imagen, sino para comparar grupos de imágenes. Sin embargo, las técnicas multivariantes se basan en una observación de todos los voxels en un único escaneo y requiere un mayor número de muestras disponibles que el de características. Por lo tanto, el clasificador debe tener en cuenta los datos de imagen de todo el cerebro. Esta metodología reporta el conocido problema de pequeño tamaño de muestra que es muy común en los estudios de medicina nuclear, ya que el número de imágenes es limitado. En este contexto de trabajo y con el objetivo claro de resolver el problema de dimensionalidad, varias técnicas propuestos para analizar las sMRI se utilizan en esta trabajo con el fin de extraer las características relevantes teniendo en cuenta la distribución de la materia gris (GM, del inglés Gray matter) y los tejidos de la sustancia blanca (WM, del inglés White matter) en la imagen del cerebro.

11.5 Descripción y preprocesamiento de base de datos

11.5.1 Base de datos utilizada: ADNI-MRI

Todos los sistemas desarrollados en esta tesis de doctorado han sido validados por un conjunto de datos de sIRM que se adquirió para el estudio de la EA. Este conjunto

de datos está etiquetado por expertos. Las etiquetas han sido asignadas para distinguir entre imágenes de pacientes con EA (en diferentes etapas de la enfermedad), imágenes de personas con deterioro cognitivo leve (MCI) que podrían conducir al Alzheimer (puede considerarse una etapa temprana de la enfermedad) y imágenes de personas sin la enfermedad (NC, sujetos de control normal). El conjunto de datos de sMRI se obtuvo de la base de datos de la Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (ADNI) (ver su página web `www.adni-info.org`).

Criterios de etiquetado

Los criterios de selección seguidos para aceptar a los participantes en el proyecto ADNI se basan en una serie de entrevistas y pruebas realizadas individualmente. Los resultados de los candidatos tenían que cumplir con ciertas condiciones de admisión al proyecto. A continuación, se detallan los criterios de selección de los pacientes para cada una de las clases de interés para el estudio:

- Sujetos normales (NC): valores MMSE entre 24-30 (incluidos), CDR de 0, no deprimido, no MCI, y no dementes. El ratio de edad de los sujetos normales se asocian rara vez a la de los sujetos MCI y EA, que suele situarse por encima de los 70 años.

- Deterioro Cognitivo Leve (MCI): valores MMSE entre 24 y 30 (incluidos), debe presentar signos de péridida de memoria, tener una pérdida objetiva de memoria medida en términos de su puntuación en el testc de Wechsler Memory Scale Logical Memory II, un CDR de 0,5, ausencia de discapacidades en otros de la función cognitiva en niveles significativos, conducta normal en las actividades de la vida cotidiana, y ausencia de demencia.

- EA precoz: valores MMSE entre 20-26 incluidos, CDR de 0.5 o 1.0, y satisface el criterio NINCDS/ADRDA [46]para EA probable.

En consecuencia, la base de datos estructural MRI se dividió en 3 clases diferentes: NC, pacientes MCI y pacientes con EA. En nuestro estudio tenemos 818 imágenes de diferentes sujetos, divididos en 229 NC, 401 MCI (312 MCI estable y 86 MCI progresiva) y 188 AD. Los datos demográficos de los pacientes en la base de datos se resumen en la tabla 11.1. La información relativa a la conversión de MCI a EA se toma de los datos clínicos disponibles en ADNI.

11.5.2 Preprocesamiento de la base de datos de imágenes

Las imágenes IRM de la base de datos ADNI fueron preprocesadas, co-registradas y segmentadas utilizando el software de mapa estadístico paramétrico (SPM) y la

Table 11.1: Detalles demográficos de la base de datos (ADNI 1075-T1)

Diagnóstico	Número	Edad	Sexo (M/F)	MMSE
NC	229	75.97±5.0	119/110	29.00±1.0
MCI	401	74.85±7.4	258/143	27.01±1.8
EA	188	75.36±7.5	99/89	23.28±2.0

herramienta de morfometría basada en vóxel (VBM8). En primer lugar, el preprocesamiento se realizó mediante un campo de deformación no lineal utilizando los mapas de probabilidad de tejido proporcionados por el "International Consortium for Brain Mapping (ICBM)". En segundo lugar, los datos se registraron en el instituto neurológico de Montreal (INM) mediante transformaciones afines y se redimensionaron en $121 \times 145 \times 121$ voxels. Posteriormente, utilizando la normalización de Dartel de alta dimensión con la plantilla Dartel estándar proporcionada por VBM8, se realizó el co-registro. Finalmente, las imágenes fueron segmentadas utilizando VBM8 de nuevo, produciendo mapas de probabilidad que consisten en valores en el rango $(0, 1)$ para cada voxel. Estos valores estiman la probabilidad de cada miembro de voxel (WM, GM o CSF), aunque no se usaron mapas CSF en nuestros experimentos.

Los datos se dividen en tres grupos diferentes:

- **Group 1:** se considera un subconjunto de los datos. No se tienen en cuenta los sujetos MCI, por lo que se forman 2 imagenes representativas EA y NC.
- **Group 2:** se considera un subconjunto de los datos. No se tienen en cuenta los sujetos EA, por lo que se forman 2 imagenes representativas MCI y NC.
- **Group 3:** se considera un subconjunto de los datos. No se tienen en cuenta los sujetos NC, por lo que se forman 2 imagenes representativas MCI y EA.

11.6 Experimentos y resultados

11.6.1 Experimentos con las sMRI segmentadas

Hemos propuesto cuatro sistemas de diagnóstico asistido por ordenador (DAO) basados en diferentes técnicas de selección y de extracción de características con el fin de apoyar el diagnóstico clínico de la EA.

- El primer y el segundo sistemas de DAO propuestos se basan en la técnica de selección de características "t-test" como reducción de dimensionalidad de las imágenes, y en las técnicas de extracción de características de "PLS" y "PCA", respectivamente, para resolver el pequeño problema de tamaño de la muestra. Las características extraídas se combinan con métodos de clasificación supervisado sy basados en "SVM" para clasificar la base de datos segmentada de sIRM.

- El tercer sistema de DAO se basa en dos modelos de "ICA" propuestos para extraer las características relevantes relacionadas con EA. Por lo tanto, el primer modelo se basa en la extracción de un bajo número de componentes independientes (IC) que funcionan como vectores de características para cada imagen de tejido de cada cerebro. El segundo se basa en la extracción de rasgos altamente representativos de cada imagen cerebral media (NC, MCI y EA) de cada tejido de cerebro. A continuación, el conjunto de fuentes de IC se utilizan como variables de entrada para el paso de la clasificación de la detección temprana de la enfermedad.

- El cuarto sistema de DAO se basa en el enfoque "NNMF" para encontrar representaciones lineales reducidas de datos no negativos. A continuación, los conjuntos de datos NNMF transformados resultantes, que contienen las características relevantes reducidas, se clasifican mediante diferentes clasificadores basados en "SVM".

11.6.2 Resultados y discusión de los sistemas DAO desarrollados

11.6.2.1 Resultados y discusión del sistema DAO desarrollado en base a las técnicas PLS:

Los resultados de la clasificación se resumen en la tabla 11.2 con diferentes experimentos (imágenes GM, imágenes WM y la combinación de características extraídas de la segmentación GM y WM) utilizando la técnica de extracción de características PLS y diferentes clasificadores SVM.

- Resultado de clasificación del grupo 1

El método PLS + SVM lineal produjo tasas de precisión más altos que núcleo RBF. Además, la combinación de características extraídas de la segmentación GM y WM dió una precisión de clasificación de 88.49 % para PLS y SVM lineal (sensibilidad

Tejidos cerebrales	Kernel	Grupo 1 Acc/Sens/Spec(%)	Grupo 2 Acc/Sens/Spec(%)	Grupo 3 Acc/Sens/Spec(%)
GM	Lineal	**87.53/88.65/86.17**	**77.57/76.76/71.9**	**77.03/74.59/79.46**
	RBF	87.29/87.77/86.7	76.22/81.62/70.81	76.22/74.59/77.84
WM	Lineal	**85.61/87.34/83.51**	**80.54/79.46/81.62**	**87.03/88.65/85.41**
	RBF	84.41/85.59/82.98	81.35/76.22/80.54	85.95/85.41/86.49
GM+WM	Lineal	**88.49/91.27/85.11**	**81.89/82.16/81.62**	**85.41/87.03/83.78**
	RBF	88.49/90.39/86.17	80.27/73.51/82.7	85.41/85.95/84.86

Table 11.2: Las medidas estadísticas del rendimiento de la técnica de selección de características PLS con diferentes clasificadores SVM, para los tres grupos de muestra, y utilizando ocho componentes. (ACC=Precisión, Sens=Sensibilidad, Spec=Especificidad)

= 91.27 % y especificidad = 85.11 %) comparado con 87.53 % sólo para GM (sensibilidad = 88.65 % Y especificidad = 86.17 %) y 85.61 % sólo para WM (sensibilidad = 87.34 % y especificidad = 83.51 %). Como conclusión, la combinación de características extraídas de GM y WM aumenta la clasificación y la precisión del clasificador.

- Resultado de clasificación del grupo 2

La tarea de clasificación más difícil en relación con la base de datos ADNI es distinguir entre NC y pacientes MCI, debido a la amplia gama de las características extraídas de los pacientes MCI. Utilizando PLS y SVM lineal, la combinación de características extraídas de la segmentación GM y WM proporcionó la mayor precisión, 81.89 % (sensibilidad = 82.16 % y especificidad = 81.62 %), mientras que las características extraídas de los tejidos cerebrales GM o WM solos informaron una precisión de clasificación de 77.57 % y 80.54 % respectivamente. En general, observamos que la combinación de características extraídas de ambos tejidos cerebrales GM y WM produjo el mayor valor de precisión.

- Resultado de clasificación del grupo 3

La combinación de características extraídas de la segmentación GM y WM presentaron una precisión de clasificación de 85.41% para PLS y SVM lineal (sensibilidad = 87.03% y especificidad = 83.78%) comparado con 77.03% sólo para GM

(sensibilidad = 74.59% y especificidad = 79.46%) y 87.03% sólo para WM (sensibilidad = 88.65% y especificidad = 85.41%). Estos resultados mostraron que el cambio más importante en el cerebro ocurre más en la sustancia blanca que en los tejidos cerebrales de la materia gris [151].
Como se muestra en la tabla 11.2, la metodología desarrollada destaca que la combinación de características extraídas de las distribuciones de tejidos cerebrales GM y WM dan mejores valores de precisión, sensibilidad y especificidad que el uso de diferentes tejidos cerebrales por separado.

- Discusión

Esta metodología logra buenos valores de precisión, sensibilidad y especificidad. Por lo tanto, la metodología propuesta es un enfoque válido para desarrollar el sistema DAO para el diagnóstico precoz de la EA. La tasa de éxito del método basado en PLS alcanzó 88.49% para el grupo 1. Sin embargo, se reduce para el grupo 2 y 3 (78.92%, 85.4% respectivamente) cuando se incluyen imágenes MCI. Esto se debe, probablemente, a la alta variabilidad del patrón MCI de cada imagen. Como consecuencia, la tarea de clasificación se vuelve más difícil. Como se muestra en el resultado del grupo 2 (NC vs MCI), el resultado de la clasificación utilizando sólo tejido cerebral WM es mejor que el uso sólo de GM. Este resultado confirma el estudio anterior [152] en el que la modificación en el patrón de atrofia cerebral en el estudio temprano (MCI) de la enfermedad se produce en el tejido cerebral WM. Además, es probable que los sujetos mayores tengan anomalías estructurales de la WM causadas por leucoaraiosis u otras enfermedades [153]. Esta anomalía en el tejido cerebral WM para los pacientes con EA o MCI puede hacer que la estructura sea muy diferente de los controles normales. Por lo tanto, el resultado de la clasificación de las imágenes del cerebro en WM puede ser mejor que en GM. Además, los resultados de la clasificación para el grupo 3 (MCI vs. EA) confirman que la neurodegeneración comienza en el WM y se extiende a GM con la progresión del trastorno.
Cabe se*ñ*alar que los sistemas DAO están reproduciendo los conocimientos médicos actuales, ya que han sido entrenados con muestras etiquetadas por los médicos. Por esta razón, las medidas estadísticas presentadas en esta metodología son una estimación de cómo un sistema es capaz de reproducir un diagnóstico médico realizado por expertos. Por lo tanto, algunos posibles errores en el proceso de etiquetado pueden modificar los hiperplanos de decisión de los clasificadores, considerando específicamente que las etiquetas fueron asignadas en base a las puntuaciones obtenidas por los pacientes en pruebas cognitivas (como MMSE y CDR).

En este trabajo, hemos demostrado anteriormente que la combinación de características extraídas de GM y WM segmentación da una buena precisión de clasificación utilizando PLS. Además, el modelo PLS + Linear SVM proporciona buenos resultados de clasificación con menor tiempo computacional.

Un enfoque más interesante consiste en seleccionar sólo algunos componentes de los métodos de extracción de características, tales FDR como se describe en [155] y el error Out-Of-Bag (OOB) en [154]. Estos métodos previos consiguen que usar los primeros componentes PLS sea óptimo para fines de clasificación. En el desarrollo de nuestro sistema DAO, hemos seleccionado sólo los 8 primeros componentes PLS. Un mayor número de componentes puede empeorar los resultados de la clasificación ya que aumenta el espacio de entrada.

11.6.2.2 Resultados y discusión del sistema DAO desarrollado en base a las técnicas PCA :

La tabla 11.3 presenta los valores de precisión, sensibilidad y especificidad de los diferentes tejidos cerebrales (GM, WM, GM + WM) utilizando la técnica PCA propuesta y los diferentes clasificadores SVM (con núcleo lineal y no lineal (RBF)). Además, el rendimiento de este sistema se calculó mediante la validación cruzada de k veces con un número de pliegues igual a 10 (k = 10)

Tejidos cerebrales	Kernel	Grupo 1 Acc/Sens/Spec(%)	Grupo 2 Acc/Sens/Spec(%)	Grupo 3 Acc/Sens/Spec(%)
GM	Lineal	**85.61/89.08/81.38**	**75.41/78.38/72.43**	**72.71/72.43/72.97**
	RBF	83.93/86.26/81.38	72.97/75.68/70.27	71.08/71.89/70.27
WM	Lineal	**81.77/84.28/78.72**	**75.14/77.3/72.98**	**79.19/82.16/76.22**
	RBF	81.29/83.41/78.72	72.7/64.86/80.54	74.86/74.05/75.68
GM+WM	Lineal	**87.77/89.96/85.11**	**78.92/80/77.84**	**81.89/84.86/78.92**
	RBF	87.55/90.39/84.04	73.24/71.89/74.59	76.77/74.59/78.92

Table 11.3: Las medidas estadísticas del rendimiento de la técnica de selección de características PCA con diferentes clasificadores SVM, para los tres grupos de muestra, y utilizando ocho componentes. (ACC=Precisión, Sens=Sensibilidad, Spec=Especificidad)

- Resultado de clasificación del grupo 1

Los resultados de clasificación aplicada al *grupo1*, que consistió en distinguir entre sujetos NC y EA, muestran un mayor valor de la tasa de precisión. La combinación

de características extraídas de la segmentación de GM y WM alcanzó una precisión de clasificación de 87.77% para PCA y SVM lineal (sensibilidad 89.96% y especificidad 85.11%) en comparación con 85.61% sólo para GM (sensibilidad 89.08% y especificidad 81.38%) y 81.77% sólo para WM (sensibilidad 84.28% y especificidad 78.72%). Como conclusión, la combinación de características extraídas de las distribuciones de ambos tejidos GM y WM aumenta la clasificación y la precisión del clasificador.

- Resultado de clasificación del grupo 2

La tarea de clasificación más difícil en este trabajo es distinguir entre sujetos NC y sujetos en las asíntotas tempranas de la enfermedad (sujetos MCI), debido a la amplia gama abarcada por las características extraídas de MCI. Utilizando PCA y SVM lineal, la combinación de características extraídas de la segmentación GM y WM proporcionó la mayor precisión, 78.92% (sensibilidad 80% y especificidad 77.84%), mientras que las características extraídas de GM o WM solo proporcionaron una precisión de clasificación de 75.41% y 75.14%, respectivamente. En general, observamos que la combinación de características extraídas de GM y WM produjo el mayor valor de precisión.

- Resultado de clasificación del grupo 3

La tabla 11.3 presenta los resultados de clasificación obtenidos en el último experimento, que consistió en distinguir entre sujetos MCI y EA usando diferentes clasificadores SVM. La combinación de características extraídas de la segmentación de GM y WM tiene un valor de precisión de clasificación del 81.89% (sensibilidad 84.86% y especificidad 78.92 %) comparado con 72.71% sólo para imágenes cerebrales de tejidos GM (sensibilidad 72.43% y especificidad 72.97%) y 79.19% para WM sólo (sensibilidad 82.16% y especificidad 76.22%). Como se muestra en la tabla 11.3, el método PCA analizado en este trabajo resalta que la combinación de características extraídas de las distribuciones de tejidos GM y WM proporcionan una mayor precisión, sensibilidad y especificidad que el uso de diferentes tejidos cerebrales por separado. Como resultado, la combinación de las diferentes características extraídas de con méthodo de clasificación ambos tejidos cerebrales (GM y WM) representa una técnica válida para realizar un sistema DAO para EA. Como se muestra en la sección anterior, el PCA es un enfoque válido para desarrollar un sistema DAO para EA. Además, el sistema DAO desarrollado consigue un buen valor de precisión, sensibilidad y especificidad. La tasa de éxito del método basado en PCA alcanzó 87.77% para el grupo1. Sin embargo, se redujo para el grupo 2 y 3 (78.92% y 81.89%, respectivamente) cuando incluyeron imágenes MCI. Esto se debe probablemente a la alta variabilidad del patrón MCI de cada imagen. Como consecuencia, la tarea de clasificación se vuelve más difícil.

11.6.2.3 Resultados y discusión del sistema DAO desarrollado basado en las técnicas ICA:

Se realizaron dos metodologías diferentes con la técnica ICA y las imágenes segmentarias del cerebro sMRI:

- **Método I**: se basa en una combinación entre la técnica de selección de características "t-test" y la técnica de extracción de características ICA.
- **Método II**: se basa únicamente en la técnica de extracción de características ICA de la imagen media de cada clase.

1. Método I

Tejidos cerebrales	Kernel	Grupo 1 Acc/Sens/Spec(%)	Grupo 2 Acc/Sens/Spec(%)	Grupo 3 Acc/Sens/Spec(%)
GM	Lineal	**84.65/86.46/82.45**	**69.46/69.03/69.96**	**69.19/70.27/68.11**
	RBF	82.97/83.41/82.45	68.38/68.65/68.11	64.05/61.62/66.49
WM	Lineal	**70.26/72.93/67.02**	**63.51/63.24/63.78**	**59.46/62.16/56.76**
	RBF	68.82/68.56/69.15	61.08/56.22/65.95	53.51/51.89/55.14
GM+WM	Lineal	**86.37/88.34/83.98**	**70.19/72.89/67.49**	**69.83/73.43/66.24**
	RBF	78.66/75.11/82.98	65.13/61.08/69.19	64.32/51.89/76.76

Table 11.4: Las medidas estadísticas del rendimiento de la técnica de selección de características ICA con diferentes clasificadores SVM, para los tres grupos de muestra, y utilizando ocho componentes. (ACC=Precisión, Sens=Sensibilidad, Spec=Especificidad)

Los resultados de la clasificación se resumen en la tabla 11.4 con diferentes experimentos (imágenes GM, imágenes WM y la combinación de características extraídas de los tejidos cerebrales GM y WM) utilizando el método de extracción de características ICA y diferentes clasificadores SVM. Puede observarse en esta tabla que el valor de precisión más alto obtenido para el grupo 1 se reduce en el grupo 2 y 3 cuando se incluyen imágenes MCI. Esta disminución de las métricas de rendimiento puede explicarse por la alta variabilidad del patrón MCI de cada imagen. En segundo lugar, se puede concluir a partir de las medidas métricas de rendimiento que

el clasificador SVM lineal produce una tasa de exactitud más alta con el enfoque de extracción de características ICA que SVM-RBF. Por lo tanto, el SVM lineal podría ser la mejor técnica para distinguir los pacientes con EA y MCI de NC. En tercer lugar, es notable que, al utilizar la combinación de caracteristicas extraidas de las imágenes GM y WM como características de entrada y posteriormente la transformación con ICA junto con SVM lineal, aumenta la precisión del clasificador y por lo tanto a*ñ*ade robustez a nuestro sistema para distinguir EA y MCI de sujetos NC.

2. Método II

Table 11.5: Medidas de rendimiento estadístico del sistema DAO propuesto basado en el modelo ICA. (ACC=Precisión, Sens=Sensibilidad, Spec=Especificidad, PL=Probabilidad positiva, NL=Probabilidad negativa)

Tipo de grupos	Tejidos cerebrales	Kernel function	Precisión	Sensibilidad	Especificidad	PL	NL
	GM	SVM-Lin	**0.881**	**0.909**	**0.851**	**6.101**	**0.106**
		SVM-RBF	**0.895**	**0.924**	**0.866**	**6.895**	**0.087**
NC.vs.EA	WM	SVM-Lin	**0.786**	**0.786**	**0.786**	**3.673**	**0.276**
		SVM-RBF	**0.799**	**0.796**	**0.802**	**4.021**	**0.254**
	GM+WM	SVM-Lin	**0.872**	**0.909**	**0.845**	**5.864**	**0.107**
		SVM-RBF	**0.866**	**0.887**	**0.844**	**5.685**	**0.134**
	GM	SVM-Lin	**0.786**	**0.813**	**0.761**	**3.402**	**0.246**
		SVM-RBF	**0.796**	**0.829**	**0.765**	**3.527**	**0.223**
NC.vs.MCI	WM	SVM-Lin	**0.739**	**0.781**	**0.696**	**2.569**	**0.315**
		SVM-RBF	**0.741**	**0.754**	**0.731**	**2.803**	**0.336**
	GM+WM	SVM-Lin	**0.776**	**0.813**	**0.739**	**3.115**	**0.253**
		SVM-RBF	**0.783**	**0.831**	**0.741**	**3.208**	**0.228**
	GM	SVM-Lin	**0.858**	**0.856**	**0.861**	**5.158**	**0.167**
		SVM-RBF	**0.855**	**0.882**	**0.831**	**5.218**	**0.142**
MCI.vs.EA	WM	SVM-Lin	**0.813**	**0.841**	**0.786**	**3.929**	**0.202**
		SVM-RBF	**0.831**	**0.866**	**0.796**	**4.245**	**0.168**
	GM+WM	SVM-Lin	**0.864**	**0.866**	**0.861**	**6.231**	**0.156**
		SVM-RBF	**0.858**	**0.876**	**0.839**	**5.441**	**0.147**

Se realizaron tres experimentos diferentes considerando los tres grupos diferentes para construir el subespacio representativo **S** descrito en la ecuación 8.4. Además, la dimensión del vector de características (m) se determinó por el número de las diferentes clases de vectores medios.

La tabla 11.6.2.3 muestra los resultados de la clasificación en el conjunto retenido para el caso de $m = 2$ (considerando vectores medios NC y EA, NC y vectores medios MCI o vectores medios MCI y EA), una vez que los pasos descritos, proyección, ICA se ha aplicado a los datos brutos.

Todos los experimentos representan una alta compresión de los datos de imagen cerebral a un peque*ñ*o número de características ($m = 2$ características).

Este método propuesto representa una alta compresión de la gran cantidad de información (voxels) contenida en la base de datos de imágenes cerebrales a un peque*ñ*o número de características.

Los resultados resumidos en la tabla 11.6.2.3 revelan que la idea de usar SVM con núcleo no lineal y un peque*ñ*o número de características genera mejores resultados que el núcleo lineal. Estos resultados obtenidos confirman las razones teóricas en [23, 129].

En cuanto a los resultados del clasificador SVM para los tres grupos con los diferentes tejidos cerebrales, se puede observar que la separación de las imágenes de tejido GM en dos clases es más eficiente. Esto podría justificarse de acuerdo con la existencia de la información más relevante que se basa en el progreso de la EA en el tejido cerebral GM.
Sin embargo, la clasificación SVM también proporciona buenos resultados con las imágenes WM. Ello significa que, incluso en la fase temprana de la enfermedad, las imágenes de WM contienen características discriminantes de EA. Por esta razón, es mejor tener consideración la distribución del tejido cerebral en el análisis de la EA en la etapa temprana.

11.6.2.4 Resultado y discusión del sistema DAO desarrollado en base de las técnicas NNMF:

En esta sección se presentan los resultados experimentales de la evaluación de los DAO, junto con sus variantes. El systema DAO basado en NNMF + SVM se desarrolla con un clasificador SVM lineal y no lineal (RBF). Así, los resultados de clasificación se resumen en la tabla 11.6 con diferentes experimentos (imágenes GM, imágenes WM y la combinación de característica extraídas de la segmentación GM y WM). De acuerdo con esta tabla, los mejores resultados en la clasificación se observan con el kernel SVM lineal. Esta metodología propuesta utilizando NNMF como técnica de extracción de características produce valores máximos de precisión = 89.04%, sensibilidad = 88.23% y especificidad = 89.83%.

Table 11.6: Medidas estadísticas de los resultados de la técnica NNMF + SVM con los diferentes "kernel". (ACC=Precisión, Sens=Sensibilidad, Spec=Especificidad)

	Lineal				RBF		
Grupos	Tejidos cerebrales	Acc	Sens	Spec	ACC	Sens	Spec
	GM	85.1064%	87.766%	82.4468%	85.9043%	88.8298%	82.9787%
NC. vs. EA	WM	82.9787%	81.9149%	84.0426%	81.383%	83.5106%	97.2553%
	(GM+WM)	**89.0391%**	**88.234%**	**89.8298%**	**86.1702%**	**83.5106%**	**88.8298%**
	GM	73.6702%	75%	72.34%	73.1383%	76.0638%	70.2128%
NC. vs. MCI	WM	76.3298%	79.7872%	72.8723%	71.5426%	72.8723%	70.2128%
	(GM+WM)	**80.0532%**	**81.9149%**	**78.1915%**	**74.4681%**	**76.0638%**	**72.8723%**
	GM	71.8085%	71.2766%	72.3404%	66.2234%	64.3617%	68.0851%
MCI. vs. EA	WM	75.7979%	76.0638%	75.5319%	75.266%	76.0638%	74.4681%
	(GM+WM)	**79.5213%**	**79.2553%**	**79.7872%**	**73.1383%**	**70.2128%**	**76.0638%**

11.7 Conclusiones

Las diversas contribuciones de esta tesis han sido probadas en una serie de experimentos, de las cuales se obtienen las siguientes conclusiones:

- Hemos propuesto varios enfoques que se basan en la la técnica de selección de características y la técnica de extracción de características para solucionar el *Small Sample Size problem* en bases de datos de neuroimagen.

- Los sistemas DAO desarrollados que se presentan en los capítulos 6, 7, 8 y 9 alcanzan un alto rendimiento en el diagnóstico diferencial de enfermos y controles, particularmente cuando se aplican a las bases de datos de Alzheimer. El mayor reto es utilizar estos sistemas para estudiar la progresión de enfermedades neurodegenerativas, y en particular, la conversión de Mild Cognitive Impairment (MCI) a demencia.

- Las técnicas de descomposición de imagen utilizadas son capaces de reducir significativamente la carga computacional cuando analizamos imágenes médicas, a la vez que mantienen o incluso incrementan el rendimiento de los sistemas DAO propuestos, que han demonstrado sobradamente su capacidad de discriminación en modalidades estructurales como MRI.

Part V

Appendix

Appendix A

Statistical Parametric Mapping

Statistical Parametric Mapping refers to the construction and assessment of spatially extended statistical processes used to test hypotheses about imaging data [18, 189]. These ideas have been instantiated in software that is called SPM [190]. This software is developed by the Wellcome Trust Centre Centre for Neuroimaging data [191]. This chapter shows an overview of SPM and its possible application to CAD systems.

A.1 Statistical Parametric Mapping

SPM is a statistical technique created by Friston [18, 189] for examining the differences in brain activity recorded using neuroimaging data such as MRI. This software is a complement tool in MATLAB [119]. In generally, it is used to identify functionally specialized brain responses and it is the most prevalent approach to characterizing the functional anatomy and the disease-related changes.
Since 1991, the College London University (UCL) provides an implementation of SPM as free software. Nowadays, it has become a reference in medical image analysis thanks to the great flexibility in designing experiments that can be performed. Thus, SPM is used in many different departments such as the department of neurology, radiology, nuclear medicine, behavior and cognitive sciences, bio-statistics, and biomedical physics worldwide [189].
In its version (SPM8), it allows to analyzing MRI images, positron emission tomography (PET) images, single photon emission computed tomography (SPECT) images, electroencephalography (EEG) and magnetoencephalography (MEG) images. Therefore, the analysis which are performed by SPM are univariate as its tests are performed voxel to voxel.
A statistical study in SPM consists of two steps: analysis and statistical inference. In addition, all images employed in the statistical study should be firstly preprocessed.

A.1.1 Preprocessing step in SPM

The preprocessing step is a fundamental step in CAD systems as it ensures that the different images are comparable to each other. This step is usually applied after the acquisition and the reconstruction of the images. The preprocessing step is divided in several processes, such as, the realign, the spatial normalization, the intensity normalization and the segmentation. However, the number and the type of procedures to follow in preprocessing step differ from one modality to another. Therefore, the normalization and the segmentation processes are used in this PhD work.

A.1.1.1 Realign

The realignment step is only applied when there are several images of the same subject. It estimated the difference of the position between the different images due to the different placements of the head in the acquisition step. In order to compensate the difference, a rotation and translation tasks are applied to these images [192].
The movement of patients could be related to the work carried out at the acquisi-

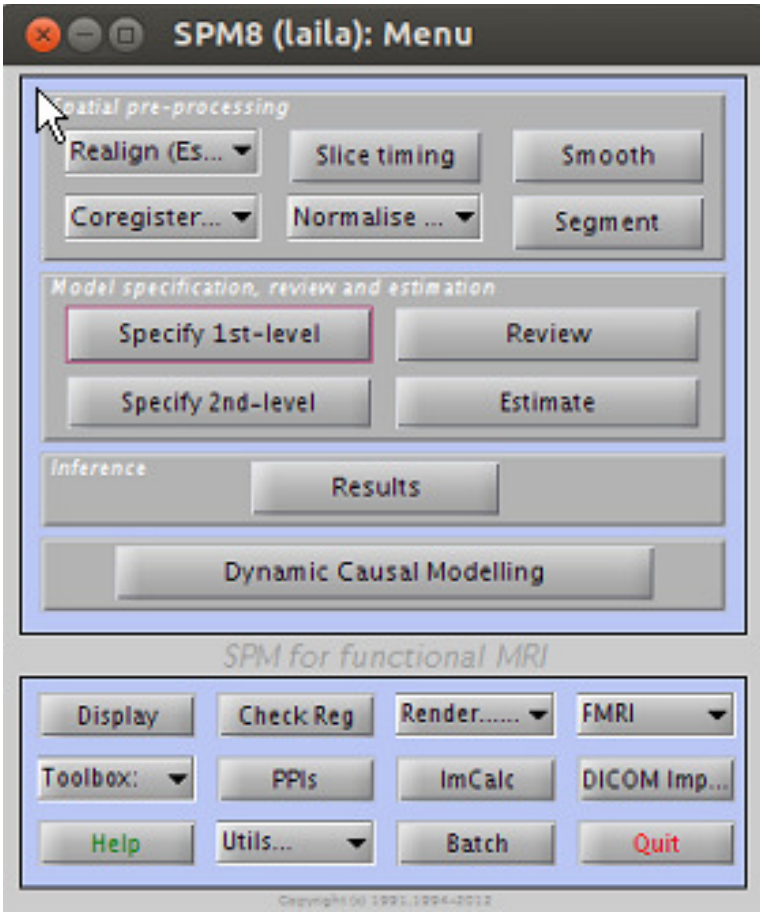


Figure A.1: The main window of SPM8 for analyzing MRI images.

tion time. For this reason, it is interesting to include the estimate of movement as a variable in the statistical analysis.
In the case of the images are from different patients, there will also be differences among each other in terms of shape and size of the brain. These differences must be corrected by the spatial normalization technique.

A.1.1.2 Spatial normalization or registration

The anatomy of every subject's brain is slightly different in shape and size. In order to compare brain images of different subjects, it is necessary to eliminate these particularities between them and transform the images so that the subsequent group analysis or comparison can be performed. Thus, the source images are mapped from their current space to a template that works as a common anatomical reference [193, 194] (See figure A.2). This procedure is known as spatial normalization or registration. After this process, each cerebral region of each subject occupies a

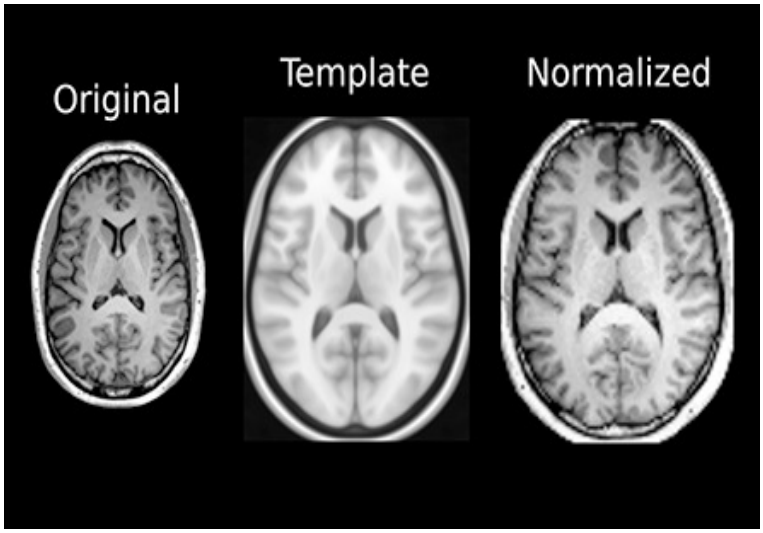


Figure A.2: An example of normalized brain image.

standard space. The Montreal Neurological Institute (MNI) space is the most widely used space for brain registration and was recently adopted by the International Consortium for Brain Mapping (ICBM) as its standard template [195]. It defines a standard three-dimensional coordinate system which is used to map the location of brain structures independently of the size and shape of each subject's brain.
The spatial normalization allows the comparison voxel-by-voxel of images. Thus, it can localize the regions of interest (ROI) related to the disease.
SPM does not cheek that the normalization is correct or not. For this reason, a visually cheek is necessary to show if the template image and the analyzed image have the same size and the same shape or not. Therefore, it is important to note that the spatial normalization equalizes the size and shape of the images under study. However, it does not eliminate the differences between them due to the metabolic characteristics of each subject.

A.1.1.3 Spatial smoothing filter

Spatial filter is a process by which the voxels are averaged with its neighbors yielding a smoothing effect whose intensity depends on a parameter called FWHM (Full Width at Half Maximum). This parameter is measured in units of space and it is usually given a value equal to three times the size of the voxel.
This preprocessing step has two principal objectives. The first one is that the spatial filter will enhance the signal-to-noise ratio (SNR) as it eliminates the image noise. The second one, it ensures that the differences between patients are given higher than a voxel so that the differences are significant ensuring sizes.
Figure A.3 presents an example of smoothed sMRI brain images.

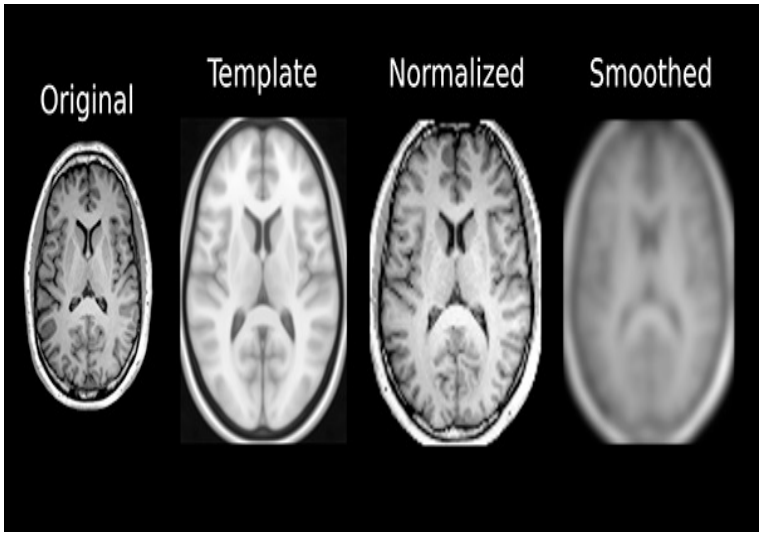


Figure A.3: An example of smoothed brain image.

A.1.1.4 Intensity normalization

The average brain metabolism varies significantly between different patients and even in the same patient at different times. For this reason, an intensity normalization process is required before to perform the statistical analysis on them. The intensity normalization process can be done in several ways:

- **Scaling:** it consists of multiplying the intensity of each image by a factor to equalize the average intensity of all the images. This approach is the simplest and the most suitable for the imaging studies coming from different patients.

- **Analysis of covariance (ANCOVA):** it consists of including the average intensity of each image in the statistical model in order to take into account the difference of intensity during the analysis. This approach performs the scaling additive which is better than the proportional scaling when the variance between the different images is low (as it is the case of one patient studies).

These two procedures are implemented in SPM but there are not suitable when the effects of interest affect sufficiently large areas to influence the average intensity value of the image. In this case, it is preferable to use other approaches such as those described below:

- **Normalization to the maximum** It consists of a division by a constant parameter. This later is often estimated as the average value of the 95th bin of the

histogram of the image, that is, the average of the 5% higher-intensity values [81, 193].

- **Normalization based on regions** This approach is similar to the proportional scaling. It scales images by equalizing the average values of the regions not affected by the pathology. Its main drawback is the need to identify and delineate the unaffected regions.

A.1.1.5 Segmentation

Mostly of structural MRI images, involve a series of algorithms aimed to construct maps of the distribution of different tissues. The general segmented approach is to separate the image in three different maps containing grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) [196–198].

In SPM, the segmentation process uses an expectation-maximization (EM) algorithm to obtain the parameters corresponding to a mixture of Gaussians that represents the different brain tissues. Afterwards, an affine transformation is applied using tissue probability maps that are in the ICBM/MNI space to extract the GM, WM and CSF maps.
Figure A.4 shows an orthogonal views of original and segmented medical image modalities.

A.1.2 Statistical analysis

SPM can perform several statistical tests, such as regression, Student t-test, F-test and analysis of variance (Anova) including variables and allowing the modeling of interactions between them [18]. These types of tests can be grouped into a general model known as the General Linear Model (GLM) which is based on two concepts: the design matrix and contrasts [119]. GLM is used by SPM to perform the mathematical calculations. Importantly, SPM did not looking for directly affect the images but only to accept or reject a hypothesis defined a priori. Therefore, a SPM study must have a hypothesis about the effects that produce the observed behavior. This hypothesis is known as the null hypothesis and it is represented by the design matrix.
Statistical studies that are performed by SPM can be divided into two types:

- Parametric or factorial studies.

- Categorical or subtractive studies.

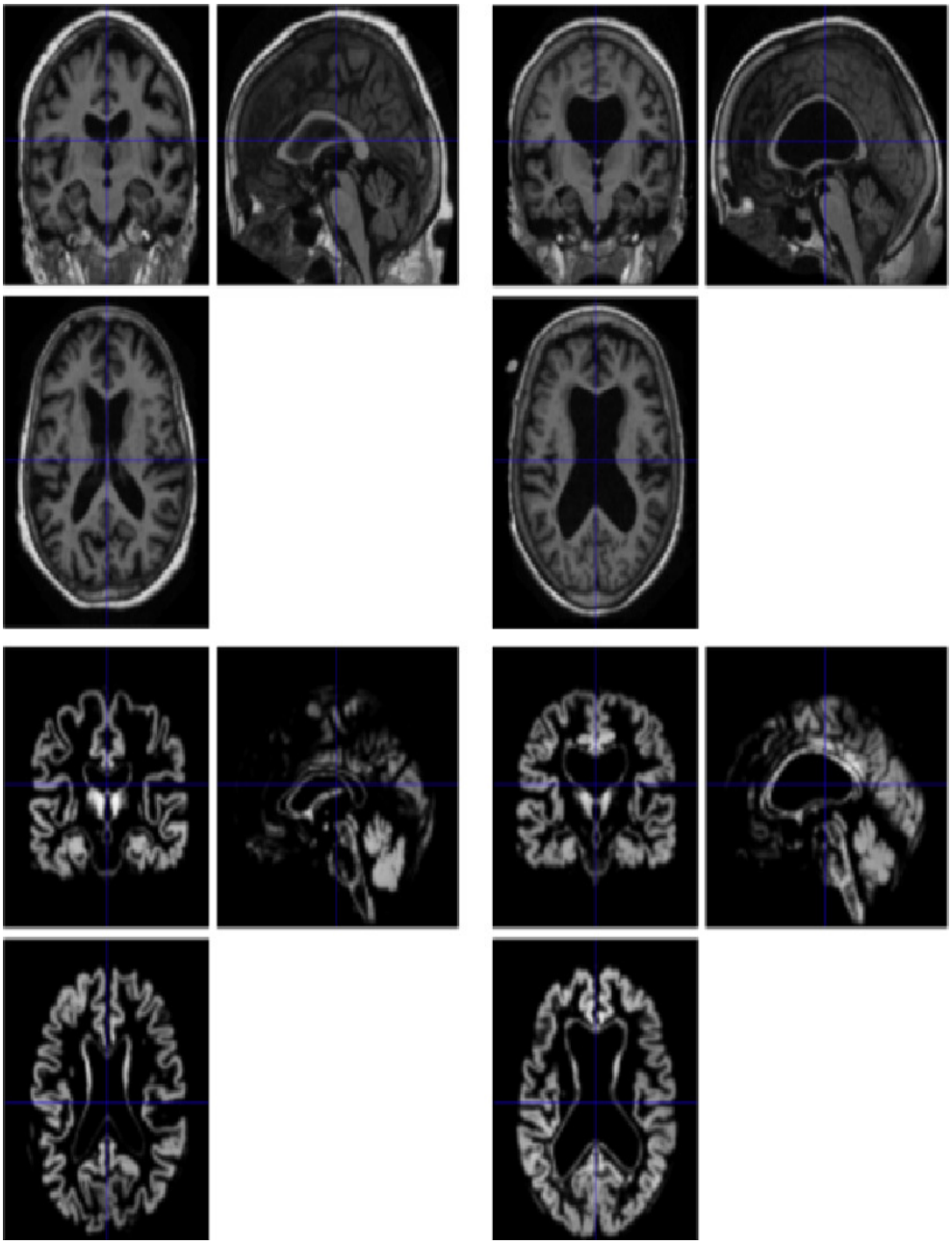


Figure A.4: Schematic illustration of different classes (AD images in the right, healthy images in the left). Orthogonal views of original and segmented medical image modalities.

The first type of studies analyzes the relationship between all images and a defined parameter, such as the age of patients or their cognitive test result.
The second type of studies are used to highlight the differences between groups defined by the categorical variables. For example, the hypothesis that age influences

the metabolism brain can be accepted or rejected by a parametric test. This hypothesis give a result in a design matrix with a row for each image included in the study and a column for the age of each patient.
Once the model are established, SPM can now automatically estimate the contribution of each effect separately. This latter allows to differentiate between effects of "interest" (for example; the effect of group) and the correctors (for example; the effect of age in parametric studies) as well as differences between the means of the two factors. By GLM, this task is done by the definition of a "contrast" which is defined as a vector.
The length of this vector is equal to the number of effects included in the design matrix, so that each effect is weighted by its corresponding element. If the effect is corrector then it is weighted with a zero vector in contrast.

In the case that the effect is parametric, the contrast determines if the desired correlation is positive, by a "1", or negative by "-1", in the effect corresponding to that position in the vector contrast. However, for categorical effects, the contrasts must comply an important condition: the sum of all the weights on the contrast in the columns of categorical effects should be equal to zero.
Finally, SPM performs the statistical test (a t-test or F-test), described by the design matrix and the contrast in all image voxels independently. Thus, the obtained result is an image whose value in each voxel is the result of statistical test and it called the *statistical parametric map*.

Bibliography

[1] R. Karen and L. Simon. The dementias. *The Lancet*, 360(9347):1759–1566, 2002.

[2] A. Wimo, L. Jonsson, J. Bond, M. Prince, and B. Winblad. The worldwide economic impact of dementia. *Alzheimers Dementia*, 9(1):1–11, 2013.

[3] PD. Meek, K. McKeithan, and G. T. Schumock. Economic considerations in Alzheimer's disease. *Pharmacotherapy*, 18(2 Pt 2):68–73, 1998.

[4] K. Blennow, M. J. de Leon, and H. Zetterberg. Alzheimer disease. *Lancet*, 368(9533):387–403, 2006.

[5] S. C. Waring and R. N. Rosenberg. Genome-wide association studies in Alzheimer disease. *Archives of neurology*, 65(3):329–334, 2008.

[6] D. B. Carr, A. Goate, D. Phil, and J. C. Morris. Current concepts in the pathogenesis of Alzheimer's disease. *The American Journal of Medicine*, 103(3A):3S–10S, 1997.

[7] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dementia*, 3(3):186–191, 2007.

[8] P. Martin, A. Wimo, M. Guerchet, Y. T. Ali. G. C, Wu, and M. Prina. *World Alzheimer Report 2015, the Global Impact of Dementia: an analysis of prevalence, incidence, cost and trends*. Alzheimer's Disease International (ADI), 2015.

[9] R. Tarawneh and D. M. Holtzman. The clinical problem of symptomatic Alzheimer disease and mild cognitive impairment. *Cold Spring Harbor Perspectives in Medicine*, 2(5), 2012.

[10] P. Sharma, S. Malik, S. Sehgal, and J. Pruthi. Computer aided diagnosis based on medical image processing and artificial intelligence methods. *International Journal of Information and Computation Technology*, 3(9):887–982, 2013.

[11] H. H. Zeng, J. Y.and Ye, S. X. Yang, R. C. Jin, Q. L. Huang, Y. C. Wei, S. G. Huang, J. Z. Wang, B. Q.and Ye, and J. Y. Qin. Clinical application of a novel computer-aided detection system based on three-dimensional CT images on pulmonary nodule. *International Journal of Clinical and Experimental Medicine*, 8(9):16077–16082, 2015.

[12] G. Zhang, P. Yan, H. Zhao, and X. Zhang. A computer aided diagnosis system in mammography using artificial neural networks. In *International Conference on BioMedical Engineering and Informatics*. IEEE, 2008.

[13] H. Fujita, Y. Uchiyama, T. Nakagawa, D. Fukuoka, Y. Hatanaka, T. Hara, G. N. Lee, Y. Hayashi, Y. Ikedo, X. Gao, and X. Zhou. Computer-aided diagnosis: the emerging of three CAD systems induced by japanese health care needs. *Computer Methods and Programs in Biomedicine*, 92(3):238–248, 2008.

[14] F. Segovia. *Análisis de imágenes funcionales cerebrales mediante modelos de mezcla de gaussianas y mínimos cuadrados parciales para el diagnóstico de alteraciones neurológicas.* PhD thesis, Universidad de Granada, 2010. Supervised by J. M. Górriz, and J. Ramírez.

[15] D. Salas-González, J. M. Górriz, J. Ramírez, M. López, I. Illán, F. Segovia, and C. G. Puntonet. Computer aided diagnosis of alzheimer's disease using support vector machines and classification trees. *Physics in Medicine and Biology*, 55(10):2807–2817, 2010.

[16] R. Chaves, J. Ramírez, J. M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, and F. Segovia. SVM-based computer-aided diagnosis of the Alzheimer's disease using T-Test NMSE feature selection with feature correlation weighting. *Neuroscience Letters*, 461(3):293–297, 2009.

[17] I. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, F. Segovia, R. Chaves, M. Gómez-Rio, and C. G. Puntonet. 18f-fdg {PET} imaging analysis for computer aided Alzheimer's diagnosis. *Information Sciences*, 181(4):903–916, 2011.

[18] K. J. Friston, J. Ashburner, S. J. Kiebel, T. E. Nichols, and W. D. Penny. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. 2007.

[19] 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. In *Medical Image Com- puting and Computer-Assisted Intervention - MICCAI*, volume 3217, pages 654–662, 2004.

[20] J. Stoeckel, G. Malandain, O. Migneco, E. M. KoulibOja, P. Robert, N. Ayache, and J. Darcourt. Classification of SPECT images of normal subjects versus images of Alzheimer's disease patients. In *Medical Image Computing*

and Computer-Assisted Intervention - MICCAI, volume 2208, pages 666–674, 2001.

[21] G. Fung and J. Stoeckel. SVM feature selection for classification of spect images of Alzheimer's disease using spatial information. *Knowledge and Information Systems*, 11:243–258, 2007.

[22] I. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, C. G. Puntonet, and F. Segovia. Alzheimer's diagnosis using eigenbrains and support vector machines. *Electronics Letters*, 45(7):342–343, 2009.

[23] R. P. W. Duin. Classifiers in almost empty spaces. In *15th Intenational Conference on Pattern Recognition*, pages 750–756. IEEE, 2000.

[24] D. Salas González, J. M. Górriz, J. Ramírez, M. López, I. Illán, C. G. Puntonet, and M. Gómez Río. Analysis of SPECT brain images for the diagnosis of Alzheimer's disease using moments and support vector machines. *Neuroscience Letters*, 461(1):60–64, 2009.

[25] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, M. López, I. Álvarez, F. Segovia, and C. G. Puntonet. Computer aided diagnosis of Alzheimer disease using support vector machines and classification trees. In *Advances in Neuro-Information Processing*, volume 5507, pages 418–425, 2008.

[26] I. Illán. *Análisis en componentes de imágenes funcionales para la ayuda al diagnóstico de la enfermedad de Alzheimer*. PhD thesis, Universidad de Granada, 2009. Supervised by J. M. Górriz, and J. Ramírez.

[27] K. Krupa and M. Bekiesińska-Figatowska. Artifacts in magnetic resonance imaging. *Polish Journal of Radiology*, 80:93–106, 2015.

[28] A. P. Dhawan. *Medical image analysis*. IEEE Press Series in Biomedical Engineering, 2003.

[29] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk. Brain imaging in Alzheimer disease. *Cold Spring Harb Perspect Med*, 2(4), 2012.

[30] D. A. Bennett and D. A. Evans. Alzheimer's disease. *Disease-a-Month*, 38(1):1–64, 1992.

[31] K. Maurer, S. Volk, and H. Gerbaldo. Auguste D. and Alzheimer's disease. *Lancet*, 11(5):561–578, 1997.

[32] A. Alzheimer. Uber eine eigenartige erkrankung der hirnrinde. *Allgemeine Zeitschrift fur Psychiatrie und Psychisch- Gerichtliche Medizin*, 64:146–148, 1907.

[33] World Health Organization. *The world health report 2001: Alzheimer's disease*. World Health Organization, 2001.

[34] World Health Organization. *Neurological disorders: public health challenges*. World Health Organization, 2006.

[35] D. Thomas and M. D. Bird. *Early-Onset Familial Alzheimer Disease*. GeneReviews, 2012.

[36] J. B. 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, E. M. Reiman, and Alzheimer's Disease Neuroimaging Initiative. Categorical and correlational analyses of baseline fluorodeoxyglucose positron emission tomography images from the Alzheimer's disease neuroimaging initiative (ADNI). *NeuroImage*, 45(4):1107–1116, 2009.

[37] N. Filippini, A. Rao, S. Wetten, R. A. Gibson, M. Borrie, D. Guzman, A. Kertesz, I. Loy-English, J. Williams, T. Nichols, B. Whitcher, and P. M. Matthews. Anatomically-distinct genetic associations of APOE epsilon4 allele load with regional cortical atrophy in Alzheimer's disease. *NeuroImage*, 44(3):724–728, 2009.

[38] Alzheimer's society. Risk factors for dementia. Technical report, Alzheimer's society, 2016.

[39] H. F. Howard. *Atlas of Alzheimer's Disease*. Informa healthcare, 2007.

[40] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen. Mild cognitive impairment: Clinical characterization and outcome. *Archives of neurology*, 56(3):303–308, 1999.

[41] P. J. Visser, P. Scheltens, and F. R. Verhey. Do MCI criteria in drug trials accurately identify subjects with predementia Alzheimer's disease. *Neurol Neurosurg Psychiatry*, 76(10):1348–1354, 2005.

[42] K. Kantarci. Magnetic resonance markers for early diagnosis and progression of Alzheimer's disease. *Expert Rev Neurother*, 5(5):663–670, 2005.

[43] B. Winblad, K. Palmer, and M. Kivipelto. Mild cognitive impairment; beyond controversies, towards a consensus: report of the international working group on mild cognitive impairment. *Internal Medicine*, 256:240–246, 2004.

[44] J. Morris. Clinical dementia rating. *Neurology*, 43(11):2412–2424, 1993.

[45] R. C. Petersen. Early diagnosis of Alzheimer's disease: Is MCI too late? *Current Alzheimer Research*, 6(4):324–330, 2009.

[46] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work groupa under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology*, 34(7):939–944, 1984.

[47] M. Folstein, S. Folstein, and P McHugh. "mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Psychiatry Research*, 12(3):189–198, 1975.

[48] D. Molloy, E. Alemayehu, and R. Roberts. Reliability of a standardized mini-mental state examination compared with the traditional mini-mental state examination. *The American Journal of Psychiatry*, 148(1):102–105, 1991.

[49] C. P. Hughes, L. Berg, W. L. Danziger, L. A. Coben, and R. L. Martin. A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, 140(6):566–572, 1982.

[50] W. G. Rosen, R. C. Mohs, and K. L. Davis. A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry*, 141(11):1356–1364, 1984.

[51] J. Sheikh and J. Yesavage. Geriatric depression scale (GDS): Recent evidence and development of a shorter version. *The Haworth Press*, 5(1–2):165–173, 1986.

[52] B. Reisberg, S. H. Ferris, M. J. Leon, and T. Crook. The globaldeterioration scale for assessment of primary degenerative dementia. *American Journal of Psychiatry*, 139:1136–1139, 1982.

[53] J. Martí-Climent, E. Prieto, J. L. Lafuente, and J. Arbizu. Neuroimagen: Fundamentos técnicos y prácticos. *Revista Españnola de Medicina Nuclear*, 29(4):189–210, 2010.

[54] K. J. Worsley and K. J. Friston. Analysis of fMRI time-series revisited - again. *NeuroImage*, 2(3):173–181, 1995.

[55] V. L. Villemagne, C. C. Rowe, K. Macfarlane, S.and Novakovic, and C. Masters. Imaginem oblivionis: the prospects of neuroimaging for early detection of Alzheimer's disease. *Clinical Neuroscience*, 12(3):221–230, 2005.

[56] F. Khalifa. *Analysis of contrast-enhanced medical images.* PhD thesis, University of Louis ville, 2014.

[57] P. Mansfield. Snapshot magnetic resonance imaging (nobel lecture). *Angewandte Chemie International Edition in English*, 43(41):5456–5464, 2004.

[58] J. P. Hornak. *The Basics of MRI*. 2011.

[59] F. F. Asrami. *Alzheimer's Disease Classification using K - OPLS and MRI*. PhD thesis, Linköping University, 2012.

[60] Magnetic resonance imaging (MRI) and MR Angiography, 2016.

[61] C. A. Cocosco, V. Kollokian, R. K. S. Kwan, G. B. Pike, and A. C. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. *Neuroimage*, 5:425, 1997.

[62] J. F. Barrett and N. Keat. Artifacts in CT: Recognition and avoidance. *Radiographics*, 24(6):1679–1691, 2004.

[63] A. D. C. Alonso. *Improvement of MRI brain segmentation*. PhD thesis, University of Denmark, 2011.

[64] A. Simmons, S. R. Arridge, G. J. Barker, and P. S. Tofts. Segmentation of neuroanatomy in magnetic resonance images. pages 2–13, 1992.

[65] M. Reuter, H. D. Rosas, and B. Fischl. Highly accurate inverse consistent registration: a robust approach. *NeuroImage*, 53(4):1181–1196, 2010.

[66] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23(1):208–219, 2004.

[67] J. B. A. Maintz and M. A. Viergever. survey of medical image registration. *Medical image analysis*, 2(1):1–36, 1998.

[68] B. BING. *Feature Extraction and matching in CONTENT-based retrieval of functional magnetic resonance images*. PhD thesis, The State University of New Jersey, 2007.

[69] J. Mazziotta, A. Toga, P. Evans, J. Fox, K. Lancaster, R. Zilles, T. Woods, G. Paus, B. Simpson, C. Pike, L. Holmes, P. Collins, D. Thompson, M. MacDonald, T. Iacoboni, K. Schormann, N. Amunts, S. Palomero-Gallagher, and L. Geyer. A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1412):1293–1322, 2001.

[70] A. Brahim. *Brain imaging biomarker normalization, analysis and classification methods to characterize the progression of neurodegenerative diseases*. PhD thesis, University of Granada, Spain, 2016. Supervised by J. M. Górriz, and J. Ramírez.

[71] J. Talairach and P. Tournoux. *Co-Planar Stereotactic Atlas ofthe Human Brain*. 1988.

[72] T. Kapur, W. E. Grimson, W. M. Wells, and R. Kikinis. Segmentation of brain tissue from magnetic resonance images. *Medical Image Analysis*, 1(2):109–127, 1996.

[73] P. Suetens, E. Bellon, D. Vandermeulen, M. Smet, G. Marchal, and J. Nuyts. Image segmentation: methods and applications in diagnostic radiology and nuclear medicine. *European Journal of Radiology*, 17(1):14–21, 1993.

[74] L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine, M. Vaidyanathan, and L. O. Hall. MRI segmentation: methods and applications. *Magn. Reson.Imaging*, 13(3):343–368, 1995.

[75] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE Transactions on Medical Imaging*, 27(5):629–640, 2008.

[76] J. Ashburner and K. Friston. Voxel–based morphometry–the methods. *Neuroimage*, 11:805–821, 2000.

[77] R. Bellman. *Adaptive Control Processes: A guided tour*. Princeton University Press, 1961.

[78] P. Cunningham. Dimension reduction. *Technical Report UCI-CSI-2007-7*, page 24, 2007.

[79] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. Illán, M. López, F. Segovia, R. Chaves, and C. G. Puntonet. Feature selection using factor analysis for alzheimer's diagnosis using 18F-FDG PET images. *Medical Physics*, 37(11):6084–6095, 2010.

[80] I. A. Gheyas and L. S. Smith. Feature subset selection in large dimensionality domains. *Pattern Recognition*, 43(1):5–13, 2010.

[81] F. J. Martínez, J. M. Górriz, J. Ramírez, C. G. Puntonet, and D. Salas González. Computer aided diagnosis tool for Alzheimer's disease based on mann–whitney–wilcoxon u–test. *Expert Systems with Applications*, 39(10):9676–9685, 2012.

[82] F. J. Martínez, J. Górriz, J. M.and Ramírez, C. G. Puntonet, and D. Salas-González. Analysis of SPECT brain images using wilcoxon and relative entropy criteria and quadratic multivariate classifiers for the diagnosis of Alzheimer's disease. In *New Challenges on Bioinspired Applications.*, volume 6687, pages 41–48. Springer Science, 2011.

[83] R. Chaves, J. Ramírez, J. M. Górriz, I. Illán, M. Gómez, Río, C. Carnero, and the Alzheimer's Disease Neuroimaging Initiative. Effective diagnosis of alzheimer's disease by means of large margin-based methodology. *BMC Medical Informatics and Decision Making*, 12(79), 2012.

[84] I. Beheshti, H. Demirel, and Alzheimer's Disease Neuroimaging Initiative. Feature-ranking-based alzheimer's disease classification from structural MRI. *Magnetic Resonance Imaging*, 34(3):252–263, 2016.

[85] M. P. Fay and M. A. Proschan. Wilcoxon-mann-whitney or t-test? on assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys 4*, 4:1–39, 2010.

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

[87] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick. Detection of prodromal Alzheimer's disease via pattern classification of MRI. *Neurobiol Aging*, 29(4):514–523, 2008.

[88] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[89] K. Fukunaga and W. L. G. Koontz. Application of the karhunen-lo232;ve expansion to feature selection and ordering. *IEEE Transactions on Computers*, 19(4):311 – 318, 1970.

[90] A. A. Miranda, Y. A. Borgne, and G. Bontempi. New routes from minimal approximation error to principal components. *Neural Processing Letters*, 27(3):197–207, 2008.

[91] M. Turk and A. Pentland. Eigenfaces for recognition. *congnitive neuroscience*, 3(1):71–86, 1991.

[92] A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4–5):411–430, 2000.

[93] I. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, F. Segovia, R. Chaves, P. Padilla, and C. G. Puntonet. Projecting independent components of {SPECT} images for computer aided diagnosis of Alzheimer's disease. *Pattern Recognition Letters*, 31(11):1342–1347, 2010.

[94] A. Hyvärinen. Survey on independent component analysis. *Neural Computation*, 2:94–128, 1999.

[95] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

[96] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *Neural Networks*, 10(3):626–634, 1999.

[97] X. Giannakopoulos, J. Karhunen, and E. Oja. Experimental comparison of neural ICA algorithms. In *the 8th International Conference on Artificial Neural Networks (ICANN 98)*, pages 651–656, 1998.

[98] K. Varmuza and P. Filzmoser. *Introduction to Multivariate Statistical Analysis in Chemometrics*. 2009.

[99] S. Wold, H. Ruhe, H. Wold, and W. J. Dunn. The collinearity problem in linear regression: the partial least squares approach to generalized inverse. *Scientific and Statistical Computing*, 5(3):735–743, 1984.

[100] H. Wold. Nonlinear partial least squares modeling ii, spline inner relation. *Chemolab*, (1–3):71–84, 1992.

[101] W. F. Massy. *Principal components regression in exploratory statistical research*. Springer, 1965.

[102] A. Phatak, P. Rilley, and A. Penlidis. The asymptotic variance of the univariate pls estimator. *Linear Algebra and its Applications*, 354(1–3):245–253, 2002.

[103] C. Goutis. Partial least squares yields shrinkage estimators. *Annals of Statistics*, 24(2):816–824, 1996.

[104] I. Frank and J. Friedman. statistical view of some chemometrics regression tools. *Technometrics*, 35(2):109–135, 1993.

[105] T. Almoy. A simulation study on comparison of prediction methods when only a few components are relevant. *Compuatational statistics and Data analysis*, 21(1):87–107, 1996.

[106] D. V. Nguyen and D. M. Rocke. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics*, 18(1):39–50, 2002.

[107] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. 1997.

[108] M. Barker and W. S. Rayens. Partial least squares for discrimination. *Chemometrics*, 2003.

[109] B. Scholkopf and A .J. Smola. *Learning with Kernels ? Support Vector Machines, Regularization, Optimization and Beyond*. 2002.

[110] E. R. Henry and J. Hofrichter. Singular value decomposition: Application to analysis of experimental data. *Numerical Computer Methods*, 210:129–192, 1992.

[111] A. Hoskuldsson. PLS regression methods. *Chemometrics*, 2(3):211–228, 1988.

[112] S. de Jong. SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263, 1993.

[113] D. D. Lee and H. S. Seung. Learning the parts of objects by nonnegative matrix factorization. *Nature*, 401:788–791, 1999.

[114] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *In NIPS*, pages 556–562. MIT Press, 2000.

[115] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics 5*, 5(2):111–126, 1994.

[116] D. Guillamet, J. Vitriá, and B. Schiele. Introducing a weighted non-negative matrix factorization for image classification. *Pattern Recognition Letters*, 24(14):2447–2454, 2003.

[117] D. Guillamet and J. Vitriá. Evaluation of distance metrics for recognition based on non-negative matrix factorization. *Pattern Recognition Letters*, 24(9–10):1599–1605, 2003.

[118] P. Sajda, S. Du, T. R. Brown, R. Stoyanova, D. C. Shungu, X. Mao, and L. C. Parra. Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain. *IEEE Transactions on Medical Imaging*, 23(12):1453–1465, 2004.

[119] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping*, 2(4):189–210, 1994.

[120] F. Falahati, E. Westman, and A. Simmons. Multivariate data analysis and machine learning in alzheimer's disease with a focus on structural magnetic resonance imaging. *Alzheimer's Disease*, 41(3):685–708, 2014.

[121] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, MO. Habert, M. Chupin, H. Benali, O. Colliot, and Alzheimer's Disease Neuroimaging Initiative. Automatic classification of patients with Alzheimer's disease from structural MRI, a comparaison of ten methods using the ADNI database. *NeuroImage*, 56(2):766–781, 2011.

[122] J. Mour ao Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter. Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. *Neuroimage*, 28(4):980–995, 2005.

[123] Z. Wang, A. R. Childress, J. Wang, and J. A. Detre. Support vector machine learning-based fMRI data group analysis. *Neuroimage*, 36(4):1139–1151, 2007.

[124] V. N. Vapnik and A. Lerner. Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24(6):774–780, 1963.

[125] V. N. Vapnik and A. Chervonenkis. A note on class of perceptron. *Automation and Remote Control*, 25, 1964.

[126] H. Selvaraj, S. Thamarai Selvi, D. Selvathi, and L. Gewali. Brain MRI slices classification using least squares support vector machine. *Intelligent Computing In Medical Sciences and Image Processing (ICMED)*, pages 21–33, 2007.

[127] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[128] A. J. Smola and B. Schölkopf. *A Tutorial on Support Vector Regression*. NeuroCOLT2 Technical Report Series NC2-TR-1998-030, 1998.

[129] V. N. Vapnik. *Statistical Learning Theory,*. WILEY–Interscience, 1998.

[130] C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1995.

[131] B. Magnin, L. Mesrob, S. Kinkingnehun, M. Pelegrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehericy, and H. Benali. Support vector machine-based classification of Alzheimers disease from whole-brain anatomical MRI. *Neuroradiology*, 2009.

[132] F. Segovia, J. M. Górriz, J. Ramírez, and D. Salas-Gonzalez. Multiclass classification of 18 F-DMFP-PET data to assist the diagnosis of parkinsonism. In *Pattern Recognition in Neuroimaging (PRNI)*, pages 1–4, 2016.

[133] A. Ortiz, J. M. Górriz, J. Ramírez, and F. J. Martínez-Murcia. LVQ-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease. In *Pattern Recognition Letters*, volume 34, pages 1725–1733, 2013.

[134] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, 1999.

[135] M. H. Zweig and G. Campbell. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4):561–577, 1993.

[136] G. Seymour. *Predictive inference*. Chapman and Hall, 1993.

[137] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Conference of Artificial Intelligence*, volume 2, pages 1137–1143. Morgan Kaufmann, 1995.

[138] Kittler J. Devijver, P. A. *Pattern Recognition: A Statistical Approach*. Prentice Hall, 1982.

[139] E. Frank, M. Hall, L. Trigg, G. Holmes, and I. H. Witten. Data mining in bioinformatics using weka. *Bioinformatics*, 20(15):2479–2481, 2004.

[140] H. Ian, E. Witten, F. Mark, and A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)*. ELSEVIER, 2011.

[141] S. Giovanni and J. Elder. Ensemble methods in data mining: Improving accuracy through combining predictions. In *Data Mining and Knowledge Discovery*, volume 2, pages 1–126, 2010.

[142] P. Refaeilzadeh, L. Tang, and H. Liu. Cross-validation. *Encyclopedia of Database Systems*, pages 532–538, 2008.

[143] B. Efron. Estimating the error rate of a prediction rule: improve- ment on cross-validation. *American Statistical Association*, 78(382):316–331, 1983.

[144] R. Rosipal and N. Krämer. Overview and recent advances in partial least squares. In *Subspace, Latent Structure and Feature Selection*, pages 34–51. Computer Science, 2006.

[145] K. J. Worsley. An overview and some new developments in the statistical analysis of PET and fMRI data. *Human Brain Mapping*, 5(4):254–258, 1997.

[146] J. Nilsson, S. de Jong, and A. K. Smilde. Multiway calibration in 3D QSAR. *Chemometrics*, 11(6):511–524, 1997.

[147] J. S. Hulland. Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Managemen*, 20(2):195–204, 1999.

[148] N. J. Lobaugh, R. West, and A. R. McIntosh. patiotemporal analysis of ex- perimental differences in event-related potential data with partial least squares. *Psychophysiology*, 38(3):517–530, 2001.

[149] R. Higdon, N. L. Foster, R. A. Koeppe, C. S. DeCarli, W. J. Jagust, C. M. Clark, N. R. Barbas, S. E. Arnold, R. S. Turner, J. L. Heidebrink, and S. Minoshima. A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging. *Statistics in Medicine*, 23:315–326, 2004.

[150] Y. Fan, N. Batmanghelich, C. M. Clark, C. Davatzikos, and Alzheimer's Disease Neuroimaging Initiative. Spatial patterns of brain atrophy in mci patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *NeuroImage*, 39(4):1731–1743, 2008.

[151] A. Xekardaki, P. Giannakopoulos, and S. Haller. White matter changes in bipolar disorder, alzheimer disease, and mild cognitive impairment: New insights from DTI. *Aging Research*, 2011:1–10, 2011.

[152] B. T. Gold, D. K. Powell, A. H. Andersen, and C. D. Smith. Alterations in multiple measures of white matter integrity in normal women at high risk for Alzheimer's disease. *Neuroimage*, 52(4):1487–1494, 2010.

[153] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M. O. Habert, M. Chupin, H. Benali, O. Colliot, and Alzheimer's Disease Neuroimaging Initiative. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage*, 56(2):766–781, 2011.

[154] F. Segovia, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. Álvarez, I. nd López, R. Chaves, and The Alzheimer's Disease Neuroimaging Initiative. a comparative study of feature extraction methods for the diagnosis of Alzheimer's disease using the ADNI database. *Neurocomputing*, 75(1):64–71, 2012.

[155] M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, and C. G. Puntonet. Automatic system for Alzheimer's disease diagnosis using eigenbrains and bayesian classification rules. In *IWANN 2009*, volume 5517, pages 949–956. Springer, 2009.

[156] C. Metz. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8(4):283–298, 1978.

[157] I. Jolliffe. *Principal Component Analysis*. Springer, 1986.

[158] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.

[159] P. G. Spetsieris, Y. Ma, V. Dhawan, and D. Eidelberg. Differential diagnosis of parkinsonian syndromes using functional PCA-based imaging features. *NeuroImage*, 45(4):1241–1252, 2009.

[160] L. Mosconi. Brain glucose metabolism in the early and specific diagnosis of alzheimer's disease. FDG-PET studies in MCI and AD. *European Journal of Nuclear Medicine and Molecular Imaging*, 32(4):486–510, 2005.

[161] L. Mosconi, W. H. Tsui, K. Herholz, A. Pupi, A. Drzezga, G. Lucignani, E. M. Reiman, V. Holthoff, E. Kalbe, S. Sorbi, J. Diehl-Schmid, R. Perneczky, F. Clerici, R. Caselli, B. Beuthien-Baumann, A. Kurz, and S. Minoshima. Multicenter standardized 18f-fdg pet diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *Nuclear Medicine*, 49:390–398, 2008.

[162] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics, 2000.

[163] S. Gauthier, B. Reisberg, M. Zaudig, R. C. Petersen, K. Ritchie, and K. Broich. Mild cognitive impairment. *Lancet*, 367(9527), 2006.

[164] A. Drzezga, N. Lautenschlager, H. Siebner, M. Riemenschneider, F. Willoch, S. Minoshima, M. Schwaiger, and A. Kurz. Cerebral metabolic changes accompanying conversion of mild cognitive impairment into Alzheimer's disease: a PET follow-up study. *European Journal of Nuclear Medicine and Molecular Imaging*, 30(8):1104–1113, 2003.

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

[166] L. Xu, G. Pearlson, and V. D. Calhoun. Joint source based morphometry identifies linked gray and white matter group differences. *Neuroimage*, 44(3):777–789, 2009.

[167] V. D. Calhoun, R. F. Silva, T. Adal, and S. Rachakonda. Comparison of PCA approaches for very large group ICA. *NeuroImage*, 118:662–666, 2015.

[168] L. Sirovich and M. Meytlis. Symmetry, probability, and recognition in face space. In *Proceedings of the National Academy of Sciences*, volume 106, pages 6895–6899, 2009.

[169] W. Yang, R. L. Lui, J. H. Gao, S. T. Chan, T. F.and Yau, and R. A. Sperling. Independent component analysis-based classification of alzheimer's MRI data. *Neuroimage*, 24(4):775–783, 2011.

[170] G. Chetelat, B. Desgranges, V. De La Sayette, F. Vi-ader, F. Eustache, and J. C. Baron. Maping gray matter loss with voxel-based morphome- try in mild cognitive impairment. *Neuroreport*, 13(15):1939–1943, 2002.

[171] A. Convit, J. de Asis, M. J. de Leon, C. Y. Tarshish, S. De Santi, and H Rusinek. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. *Neurobiology of aging*, 21(1):19–26, 2000.

[172] B. C. Dickerson, I. Goncharova, M. P. Sullivan, C. Forchetti, R. S. Wilson, and D. A. Bennett. Mri-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. *Neurobiology of aging*, 22(5):747–754, 2001.

[173] R. J. Killiany, T. Gomez-Isla, M. Moss, R. Kikinis, T. Sandor, and F. Jolesz. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. *Annals of neurology*, 47(4):430–439, 2000.

[174] R. Filipovych and C. Davatzikos. Semi-supervised pattern classification of medical images: Application to mild cognitive impairment (MCI). *Neuroimage*, 55(3):1109–1119, 2011.

[175] E. E. Bron, M. Smits, W. M. van der Flier, H. Vrenken, F. Barkhof, and P. Scheltens. Standardized evaluation of algorithms for computer- aided diagnosis of dementia based on structural MRI: The CAD dementia challenge. *Neuroimage*, 111:562–579, 2015.

[176] M. R. Sabuncu and E. Konukoglu. Clinical prediction from structural brain mri scans: A large-scale empirical study. *Neuroinformatics*, 13(1):31–46, 2014.

[177] L. Harper, G. G. Fumagalli, F. Barkhof, P. Scheltens, J. T. O'Brien, F. Bouwman, E. J. Burton, J. D. Rohrer, N. C. Fox, G. R. Ridgway, and J. M. Schott. MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. *Brain*, 139(Pt 4):1211–1225, 2016.

[178] A. A. Willette, V. D. Calhoun, J. M. Egan, and D. pogiannis. Prognostic classification of mild cognitive impairment and alzheimer's disease: MRI independent component analysis. *Psychiatry Research:Neuroimaging*, 224(2):81–88, 2014.

[179] D. Chyzhyk, M. Grana, A. Savio, and J. Maiora. Hybrid dendritic computing with kernel-lica applied to alzheimer's disease detection in MRI. *Neurocomputing*, 75(1):72–77, 2012.

[180] L. Khedher, J. Ramírez, J. M. Górriz, and A. Brahim. Automatic classification of segmented MRI data combining independent component analysis and support vector machines. In *Innovation in Medicine and Healthcare–InMed*, volume 207 of *Studies in Health Technology and Informatics*, pages 271–279, 2014.

[181] S. Kloppel, C. M. Stonnington, C. Chu, B. Dragan-ski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack Jr, J. Ashburner, and R. S. J. Frackowiak. Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(Pt 3):681–689, 2008.

[182] D. Daniel, H. Lee, and S. Sebastian. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562. MIT Press, 2000.

[183] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, and F. Segovia. Early diagnosis of alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented mri images. *Neurocomputing*, 151(1):139–150, 2014.

[184] L. Khedher, I. Illán, J. Ramírez, J. M. Górriz, A. Brahim, and A. Meyer-Baese. Independent component analysis-support vector machine-based computer - aided diagnosis system for alzheimer's with visual support. *Neural Systems*, 27(3):1–19, 2016.

[185] L. Khedher, J. Ramírez, J. M. Górriz, and A. Brahim. Independent component analysis-based classification of alzheimer's disease from segmented mri data. In *International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC)*, volume 9107, pages 78–87, 2015.

[186] R. Swerdlow. Brain aging, Alzheimer's disease, and mitochondria. *Biochimica et Biophysica Acta*, 1812(12):1630–1639, 2011.

[187] B. Vellas, L. Froelich, and C. Sampaioc. Commentary on health economics and the value of therapy in alzheimer's disease. value therapy for Alzheimer's disease a european perspective. *Alzheimer's and Dementia*, 3(3):152–156, 2007.

[188] J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, K. Meguro, J. O'brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, Y. Stern, P. J. Visser, and P Scheltens. Research criteria for the diagnosis of Alzheimer's disease: revising the nincds-adrda criteria. *Lancet neurology*, 6(8):734–746, 2007.

[189] J. D. Gispert, J. Pascau, S. Reig, P. Garcia-Barreno, and M. Desco. Mapas de estadisticos paramétricos (SPM) en medicina nuclear. *Revista Espanola de Medicina Nuclear*, 22(1):43–53, 2003.

[190] Grandinetti. Lucio, editor. *Grid Computing: The New Frontier of High Performance Computing*. ELSEVIER, 2005.

[191] F. Guillaume and J. Marianne. *fMRI Data Analysis Using SPM*. Springer, 2013.

[192] R. P. Woods, S. R. Cherry, and J. C. Maziota. Rapid automated algorithm for aligning and reslicing PET images. *Computer assisted tomography*, 16(4):620–633, 1992.

[193] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. Illán, M. López, F. Segovia, R. Chaves, and P. Padilla. Feature selection using factor analysis for alzheimer's diagnosis using 18F-FDG PET images. *Medical Physics*, 37(11):6084–6095, 2010.

[194] J. Ashburner, P. Neelin, D. L. Collins, A. C. Evans, and K. J. Friston. incorporating prior knowledge into image registration. *Neuroimage*, 6:344–352, 1997.

[195] A. C. Evans, D. L. Collins, S. R. Mills, E. D. Brown, and R. L. Kelly. 3D statistical neuroanatomical models from 305 MRI volumes. In *Nuclear Science Symposium and Medical Imaging Conference*. IEEE, 1993.

[196] F. Segonne, A. M. Dale, D. Salat, H. K. Hahn, and B. Fischl. A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, 22(3):1060–1075, 2004.

[197] Fischl B. Dale, A. M. and M. I. Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194, 1999.

[198] 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*, 33(3):341–355, 2002.