Structural MRI-based Classification of Alzheimer's Disease

Iman Beheshti

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical and Electronic Engineering

Eastern Mediterranean University
February 2016
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Cem Tanova
Acting Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Doctor of Philosophy in Electrical and Electronic Engineering.

Prof. Dr. Hasan Demirel
Chair, Department of Electrical and Electronic Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis of the degree of Doctor of Philosophy in Electrical and Electronic Engineering.

Prof. Dr. Hasan Demirel
Supervisor

Examining Committee

1. Prof. Dr. Hasan Demirel
2. Prof. Dr. Aytül Erçil
3. Prof. Dr. Şener Uysal
4. Prof. Dr. Adnan Yazici
5. Assoc. Prof. Dr. Önsen Toygar

ABSTRACT

Alzheimer's disease (AD), an irreversible neurodegenerative dementia, occurs most frequently in older adults which gradually destroys regions of the brain that are responsible for memory, learning, thinking and behavior. By estimation, 5.3 million Americans of all ages suffered from AD in 2015. This number is expected to increase to 16 million people by 2050. AD is the only cause of death in the top 10 of Americans that cannot be cured, prevented or slowed. Presently, there is no cure for AD, but early detection may help to figure out the root of AD mechanisms and improve the quality of life for patients who suffer from AD. In recent years, analysis of neuroimaging data has attracted a lot of interest with the recent improvements for early and accurate detection of AD. Neuroimaging techniques have become an important field of research due to the progress in their acquisition, storage and management in a wide range of applications including AD detection. High accurate image-based early detection of AD could provide valuable support for clinical treatments. High-dimensional classification methods have been a major target in the field of machine learning for the automatic AD detection. One major issue of automatic AD classification is the feature-selection method from high-dimensional feature space. This study proposes novel feature selection methods for high-dimensional pattern recognition problem aimed at high accurate detection of AD, which uses the information from three dimensional magnetic resonance imaging (MRI) data extracted from the brain.

MRI-based brain data used in the present study are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). This work focuses on structural MRI data

and investigates extraction and selection of features, which are the main blocks in an automatic diagnosis detection system. In this regard, Voxel-based-morphometry (VBM) analysis of cross-sectional 3-Tesla 3D T1-weighted MRI data is utilized to perform feature extraction. VBM is an automated technique for assessment of whole brain structure with voxel-by-voxel comparisons which has been developed to analyze tissue concentrations or volumes between subject groups to distinguish degenerative diseases with dementia. The significant local differences in gray matter volumes (gray matter atrophies) based on VBM analysis are selected as 3-D volumes of interests (VOIs). Feature extraction based on the 3D voxel clusters detected by VBM on structural MRI (sMRI) and voxel values of VOIs are considered as raw features. In the feature selection stage, novel methods based on probability distribution function (PDF) and feature ranking are introduced to select most discriminative features from high-dimensional data. In the PDF-based feature selection approach, a novel statistical feature-selection process is employed, utilizing the PDF of the VOI to represent statistical patterns of the respective high-dimensional sMRI sample. PDF of the VOIs can be considered a lower-dimensional feature vector representing sMRI images. The dimensionality of the PDF-based feature vector can be adjusted by changing the number of bins of the PDF. In this regard, the Fisher Criterion is used to determine the optimal number of bins of the histogram generating the PDF. In the proposed feature ranking method, all raw features are ranked using seven different statistical measures methods, namely, statistical dependency (SD), mutual information (MI), information gain (IG), Pearson's correlation coefficient (PCC), t-test score (TS), Fisher's criterion (FC), and the Gini index (GI). These measures are indicators of class separability, therefore the features with higher scores are assumed to be more discriminative. Hence it is critical

to determine the number of top features. In the current study, to determine the number of top features, two methods namely, Fisher criterion and classification error are introduced. The Fisher Criterion between AD and HC groups is calculated for all sizes of feature vectors, where the vector size maximizing Fisher Criterion is selected as the number of top discriminative features. In a similar spirit, the estimated classification error on training set made up of the AD and HC groups is calculated. The vector size that minimizing this error is selected as the size of the top discriminative feature vector. In the classification stage, the support vector machine (SVM) classifiers with linear and non-linear kernels are employed to perform binary classification using 10 fold cross validation between patients who suffer from AD and age-matched healthy controls. Moreover, data fusion techniques are proposed to achieve higher performance in AD detection. In this regard, data fusion is introduced to improve the classification performance, by combining scores or vectors received from clusters obtained from MRI images based on the severity of gray matter atrophy in the brain. In addition, a novel data fusion approach among feature ranking methods is introduced. The results indicate that proposed approaches are reliable techniques that are highly competitive with the state-of-the-art techniques in classification of AD.

Keywords: Alzheimer's disease, Structural MRI, Voxel-based morphometry, Statistical feature extraction, Probability distribution function, Feature ranking, Fisher Criterion, classification error, Data fusion, , Support vector machine.

ÖZ

Alzheimer hastalığı (AH), geri dönüşü olmayan bir nörodejeneratif bunaklık hastalığı olup, sıklıkla yaşlı erişkinlerde beynin hafıza, öğrenme, düşünme ve davranış ile ilgili bölgelerini yavaş yavaş yok ederek ortaya çıkmaktadır. Tahminlere göre, her yaştan 5.3 milyon Amerikalı 2015 yılı itibarıyle Alzheimer hastalığından muzdariptir. Bu sayının 2050 yılında 16 milyona yükselmesi beklenmektedir. AH, tedavi edilebilir, önlenebilir ya da yavaşlatılabilir bir hastalık olmayıp Amerikalılar arasında en yüksek ilk 10 ölüm nedenleri arasında yer almaktadır. Halen, ortada AH tedavisi bulunmamakla birlikte erken teşhis AH mekanizmalarını anlamaya ve bu hastalıktan muzdarip insanların yaşam kalitesini artırmak için yardımcı olabilir. Son yıllarda, beyin görüntü verilerinin analizi ile AH'nin erken ve doğru tespiti için ortaya çıkan gelişmeler çok ilgi çekmektedir. Beyin görüntüleme tekniklerindeki gelişmeler sayesinde veri edinim, depolama ve yönetimi konuları önemli bir araştırma alanı oluşturarak AH tespiti de dahil olmak üzere geniş bir uygulama yelpazesi ortaya çıkarmaktadır. AH'nin yüksek doğrulukla görüntü tabanlı erken teşhisi, klinik tedaviler için değerli bir destek sağlayabilmektedir. Yüksek boyutlu sınıflandırma yöntemleri otomatik AH tespiti için makine öğrenme alanında önemli bir hedef olmuştur. Otomatik AH sınıflandırma yaklaşımlarında önemli bir konu da yüksek boyutlu öznitelik uzayından öznitelik-seçme yöntemidir. Bu çalışmada, üç boyutlu MR beyin verilerinden çıkarılan bilgiler kullanılarak AH'nin yüksek doğrulukla tespiti hedefiyle yüksek boyutlu tanıma problemi için yeni özellik seçim yöntemleri önerilmektedir. Bu çalışmada kullanılan Manyetik Resonans Görüntüleme (MRG) tabanlı beyin verileri Alzheimer Hastalığı Beyin Girişimi (ADNI) tarafından oluşturulmuştur.Sunulan bu çalışmada, yapısal MRG verileri

incelenerek çıkarılan ve seçilen öznitelikler otomatik teşhis algılama sisteminin temel taşları olarak çalışılmaktadır. Bu bağlamda, kesitsel 3 Tesla 3B T1 ağırlıklı MR verilerinin voksel-bazlı morfometri (VBM) analizi özellik çıkarımını gerçekleştirmek için kullanılmaktadır. VBM ile dejeneratif hastalıklar ile bunaklık hasta gruplarını ayırt etmek için doku konsantrasyonlarını veya birimleri analiz etmek mümkün olmaktadır. VBM tekniği ile konu grupları arasındaki voksel, voksel karşılaştırmalar ile tüm beyin yapısının değerlendirilmesi otomatik olarak mümkün olmaktadır. VBM analizi tabanlı gri madde hacimlerinde önemli yerel farklılıklar (gri madde körelmesi) meydana gelmekte ve bu bölgeler 3B ilgi hacimleri (VOIs) olarak seçilmektedir. Yapısal MRG ve VOI ham voksel değerleri üzerinden VBM tarafından algılanan 3B voksel kümelerine dayalı öznitelik çıkarımı yapılmaktadır. Öznitelik seçimi aşamasında olasılık dağılım fonksiyonu (PDF) ve öznitelik sıralaması tabanlı yeni yöntemler önerilmekte, yüksek boyutlu ham verilerin en ayırt edici özellikleri seçilebilmektedir. PDF tabanlı öznitelik seçimi yaklaşımında, yeni bir istatistiki öznitelik seçim süreci önerilmekte ve bu bağlamda ilgili yapısal MRG örneklerden elde edilen VOI üzerinden çıkarılan PDF seçilen yüksek boyutlu bölgenin istatistiksel örüntüsünü temsil etmek için kullanılmaktadır. VOI'lerden çıkarılan PDFler yapısal MRI görüntülerini temsil eden düşük boyutlu öznitelik vektörleri olarak kabul edilebilmektedir.

PDF tabanlı özellik vektörünün boyutu PDF bidonlarının sayısı değiştirilerek ayarlanabilmektedir. Bu bağlamda, Fisher kriteri kullanılarak PDF'i üreten histogram bidonlarının optimal sayısı belirlenebilmektedir. Önerilen öznitelik sıralama yönteminde, tüm ham öznitelikler, yedi farklı istatistiksel ölçüm yöntemleri kullanılarak sıralanabilmektedir. Bu yöntemler sırasıyla: İstatistiksel bağımlılık (SD),

karşılıklı bilgi (MI), bilgi kazancı (IG), Pearson korelasyon katsayısı (PCC), t-test puanı (TS), Fisher kriteri (FC) ve Gini indeksi (GI) olarak seçilmiştir. Bu ölçümler sınıflar arası ayrılabilirlik ölçüsünü göstermektedir. Bu nedenle ölçümlerdeki yüksek değerler kullanılan özniteliklerin daha ayrımcı olduğunu göstermektedir.Dolayısıyla en üst özniteliklerin sayısını belirlemek çok önemlidir. Bu çalışmada, en üst özniteliklerin sayısını belirlemek için iki yöntem yani Fisher kriteri ve sınıflandırma hatası önerilmektedir. AH ve sağlıklı kontol (HC) grupları arasında Fisher Kriteri, öznitelik vektörlerinin tüm boyutları için hesaplananmakta ve Fisher kriterini maksimize eden vektör boyutu en üst ayrımcı öznitelik vektör boyutu olarak seçilmektedir. Benzer bir yaklaşımla, AH ve HC gruplarından oluşan eğitim seti üzerinde sınıflandırma hatası hesaplanmaktadır. Bu hatayı minimize eden boyut, en üst ayırt edici öznitelik vektörünün boyutu olarak seçilmektedir.

Sınıflandırma aşamasında, doğrusal ve doğrusal olmayan çekirdekli destek vektör makinesi (SVM) sınıflandırıcılarının AH ve yaş uyumlu sağlıklı kontrollerden muzdarip hastalar arasında 10 kat çapraz doğrulama kullanarak ikili sınıflandırma yapmaktadır. Ayrıca, veri füzyonu teknikleri AH tespitinde daha yüksek performans elde etmek için önerilmiştir. Bu bağlamda, veri füzyonu beyinde gri madde körelme şiddetine göre MR görüntülerinden elde edilen kümelere alınan puanlar ya da vektörler birleştirerek, sınıflandırma performansını artırmak için önerilmiştir. Buna ek olarak, özellik sıralamasında yöntemler arasında yeni bir veri füzyon yaklaşımı tanıtılmıştır. Sonuçlar, önerilen tekniklerin AH sınıflandırılmasında literatürdeki alternatif teknikler ile son derece rekabetçi ve güvenilir teknikler olduğunu göstermektedir.

Anahtar Kelimeler: Alzheimer hastalığı, yapısal MRG, voksel-temelli morfometri, İstatistiksel özellik çıkarımı, olasılık dağılım fonksiyonu, öznitelik sıralaması, Fisher kriteri, sınıflandırma hatası, veri füzyonu, vektör destek makinesi.

Dedicated to

my wife who has always been supportive of me during my time at EMU

and

my parents.

ACKNOWLEDGMENT

My special thanks go to my supervisor Prof. Dr. Hasan Demirel, who has patiently and tirelessly guided me and kept me going through his encouragement and enthusiasm over the past few years.

Thanks also go to Prof. Hiroshi Matsuda from Integrative Brain Imaging Center, National Center of Neurology and Psychiatry, Tokyo, Japan, for his recommendations during pre-processing steps and providing pre-processing software.

In addition, I would like to thank Prof. Chunlan Yang from College of Life Science and Bioengineering, Beijing University of Technology, Beijing, China, for her recommendations during this research.

Last but not least, a heartfelt thank to my wife and parents, for their support and encouragement at every stage.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

3-D	Three Dimension
μ	Mean
σ	Standard deviation
λ	Eigen value
E	Error
C	Regularization
C_i	Class number
K	Fold
$K(.,.)$	Kernel function
N_{bin}	number of bins
N_{opt}	optimal number of bins
S_i	Support Vector
S_B	between-class scatter matrix
S_W	within-class scatter matrix
$\bar{x}$	Mean
ACC	Accuracy
AD	Alzheimer's Disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
ANN	Artificial Neural Network
AUC	Area Under a Curve
CAD	Computer-Aided Diagnosis
CDR	Clinical Dementia Ratio

CSF	Cerebrospinal Fluid
CT	X-ray computed Tomography
CV	Cross Validation
DARTEL	Diffeomorphic Anatomic Registration Through Exponentiated Lie algebra algorithm
DBM	Deformation-Based Morphometry
DM	Displacement Magnitude
EEG	Electroencephalography
FC	Fisher Criterion
FP	False Positive
FPR	False-Positive Ratio
FN	True Negatives
Fmri	Functional Magnetic Resonance Imaging
FEW	Family-Wise Error
FWHM	Full-Width-Half-Maximum
GI	Gini index
GR	Gradient Recalled
GM	Gray Matter
GLM	General Linear Model
H	Histogram
HC	Healthy Control
IG	Information Gain
MEG	Magneto Encephalography
MI	Mutual Information

MNI	Mosntreal Neurological Institute
MMSE	Mini Mental State Examination
MRI	Magnetic Resonance Imaging
OASIS	Open Access Series of Imaging Studies
PCA	Principal Component Analysis
PCC	Pearson's correlation Coefficient
PDF	Probability Distribution Function
PET	Positron Emission Tomography
PLS	Partial Least Squares
RBF	Radial Basis Function
ROC	Receiver Operating Characteristic
ROI	Regions of Interest
SBA	Surface-Based Morphometry
SD	Statistical Dependency
SE	Standard Error
SEN	Sensitivity
SPE	Specificity
SPECT	Single Photon Emission Computed Tomography
SPM	Statistical Parameter Mapping
sMRI	Structural Magnetic Resonance Imaging
STAND	STructural Abnormality iNDex
SVM	Support Vector Machine
TBM	Tensor Based Morphometry
TE	Echo Time

TI	Inversion Time
TN	True Negative
TP	True Positive
TR	Repetition Time
TS	t-test Score
VBM	Voxel-Based Morphometric
VoI	Volume of Interest
WM	White Matter

Chapter 1

INTRODUCTION

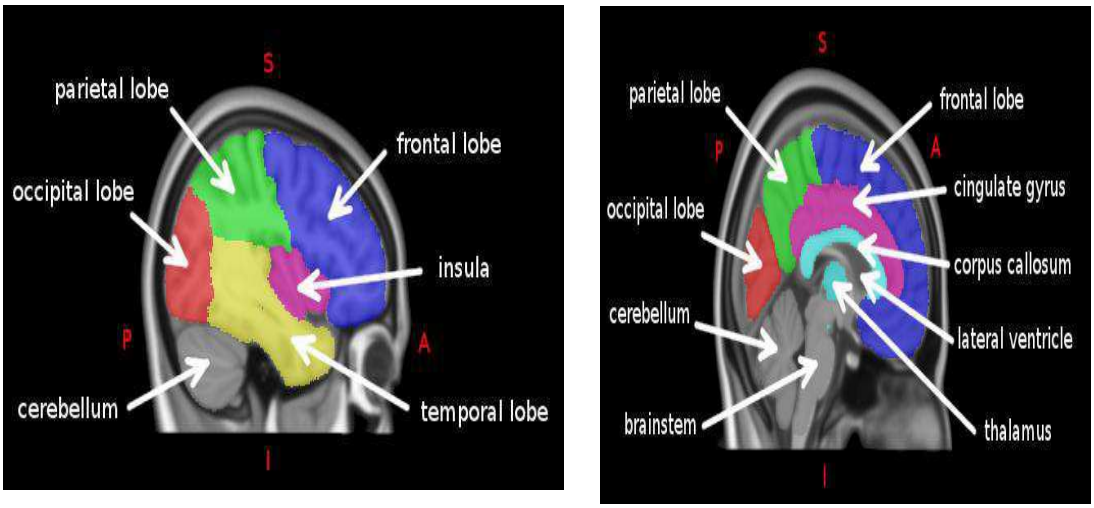
1.1 Introduction

In older adults, Alzheimer's disease (AD) is a brain disorder that gradually impairs regions of the brain that are responsible for memory, learning, and higher executive functioning (Carter, Resnick, Mallampalli, & Kalbarczyk, 2012; Seixas, Zadrozny, Laks, Conci, & Muchaluat Saade, 2014). Current estimates indicate that 5.3 million Americans of all ages will suffer from AD in 2015. This number is expected to increase to 16 million people by 2050. AD is the only disease among the top ten causes of death in Americans that cannot be cured, prevented, or slowed ("Alzheimer's Association | Alzheimer's Disease and Dementia," 2015). Presently, no cure exists for AD, but early detection may aid in determining the root of AD mechanisms and improve the quality of life for patients who suffer from AD ("Alzheimer's Association | Alzheimer's Disease and Dementia," 2015). Currently, clinical trials are investigating on development of new treatment to help patients who suffer from AD to maintain mental function and manage the behavioural symptoms. In general, changes associated with AD, occur many years before the onset of clinical symptoms such as losing memory, aggression, preoccupation with bodily functions, and apathy reclusive behavior, emotional lability, hoarding, and refusal of help. The early detection of AD may help in understanding the root of AD mechanisms as biomarkers for detection and monitoring, and also help scientists and

clinicians to develop relevant, targeted treatments. In this aim, Neuroimaging data may help to reveal markers for the early diagnosis of AD. The aim of the current research presented in this thesis is to use Neuroimaging data using machine learning methods to identify patients who suffer from AD.

1.2 Neuroanatomy

The human brain, illustrated in Figure 1.2, is composed mainly of two cerebral hemispheres, each of which is divided into four lobes: frontal, temporal, parietal and occipital. Each hemisphere includes a cortex of grey matter containing the neuronal cell bodies. The cortical surface is folded into ridges (gyri) and grooves (sulci). Other cortical regions relevant to the study of AD include the cingulate gyrus and insula. The insula is folded deep within the lateral sulcus between the frontal and temporal lobes. On the lateral surface of the brain, it is covered by the operculum, which is formed from portions of the frontal, temporal and parietal lobes.



(a) Lateral view (b) Medial view

Figure 0.1: Sagittal views of the right hemisphere of the brain, showing its gross anatomy. S: superior, I: inferior, A: anterior, P: posterior("Alzheimer's Association | Alzheimer's Disease and Dementia," 2015)

The cortex surrounds a core of white matter, consisting mainly of myelinated axons connecting the cell bodies. The largest white matter structure in the brain is the corpus callosum, a bundle of axons connecting the left and right cerebral hemispheres. Embedded within the cerebral white matter are deep grey matter structures, including the basal ganglia and thalamus. At the base of the brain, underneath the cerebral hemispheres, are the cerebellum and brainstem. The brainstem is continuous with the spinal cord. The brain is separated from the skull by three layers of tissue known as meninges: the dura, the arachnoid and the pia. To protect and support the brain, cerebrospinal fluid (CSF) fills the subarachnoid space, as well as a continuous system of four cavities known as ventricles.

1.3 Neuroimaging

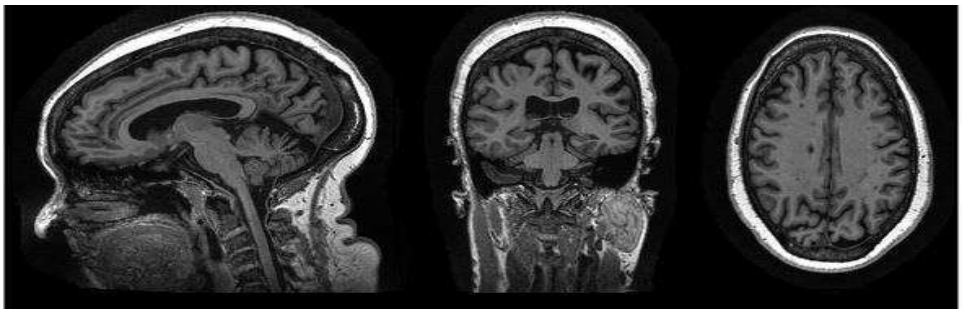
Currently, the detection of AD is based on clinical examinations and assessments of perception and behavior as indicators emerging in the later disease stages. Neuroimaging measures of structural changes and functional activities in the brain may be a good method for early detection of AD. In recent years, the analysis of neuroimaging data has attracted much interest, given the recent improvements in early and accurate detection of AD (S. Liu et al., 2014; Weiner et al., 2015) such as magnetic resonance imaging (MRI), positron emission tomography (PET) , single photon emission computed tomography (SPECT) and X-ray computed tomography (CT). Among the several available neuroimaging modalities, magnetic resonance imaging (MRI) is more widely used in AD related studies because of its excellent spatial resolution, high availability, good contrast, and the lack of a requirement for the radioactive pharmaceutical injection that is needed with positron emission tomography (PET) or single photon emission computed tomography (SPECT) (Chen, Deutsch, Satya, Liu, & Mountz, 2013; Górriz, Segovia, Ramírez, Lassl, & Salas-

Gonzalez, 2011; Gray et al., 2012; Hanyu et al., 2010). In this thesis we mainly focus on AD classification using structural MRI.

1.4 MRI biomarkers for Alzheimer's disease

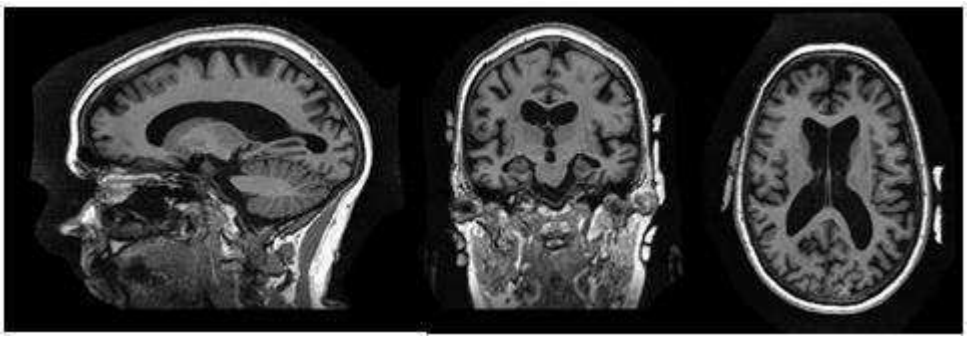
Recently, several studies have used biomarkers to classify AD based on structural MRI (Aguilar et al., 2013; I. Beheshti & Demirel, 2015b; Bron et al., 2015; M. Li, Qin, Gao, Zhu, & He, 2014; Moradi, Pepe, Gaser, Huttunen, & Tohka, 2015; Papakostas, Savio, Graña, & Kaburlasos, 2015; Westman, Muehlboeck, & Simmons, 2012; D. Zhang, Wang, Zhou, Yuan, & Shen, 2011), which can be utilized to specify brain atrophy; functional MRI (Andersen, Rayens, Liu, & Smith, 2012; Dinesh, Kumar, Vigneshwar, & Mohanraj, 2013; Fan, Resnick, Wu, & Davatzikos, 2008), which can be employed to describe hemodynamic response relevant to neural activity; diffusion tensor imaging (Graña et al., 2011; Lee, Park, & Han, 2013; Mesrob, 2012), which can be used for local microstructural characteristics of water diffusion; and functional/structural connectivity(Challis et al., 2015; Shao et al., 2012; Wee et al., 2012), which can be used to characterize neurological disorders in the whole brain at the connectivity level. In this thesis we mainly focus on AD classification using structural MRI. Atrophy measured by structural MRI is a powerful biomarker of the stage and intensity of the neurodegenerative aspect of AD pathology (Vemuri & Jack, 2010).

Figure 1.2 shows brain atrophy in AD and HC using sMRI modality.



Saggital View Coronal View Axial View

(a)



Saggital View Coronal View Axial View

(b)

Figure 1.2: The sMRI of (a) healthy individuals, and (b) AD patients with atrophy

Several studies have used structural MRI feature extraction for AD classification. These studies are variously based on morphometric methods (Huang, Yan, Jiang, & Wang, 2008; Savio et al., 2011; J. Z. J. Zhang, Yan, Huang, Yang, & Huang, 2008), region of interest (ROI)/volume of interest (VOI) (Fung & Stoeckel, 2007; Lao et al., 2004; Yanxi Liu, 2004), gray matter voxels in the automatic segmentation of images (Klöppel et al., 2008), and structural MRI measurement of the hippocampus and the medial temporal lobe (Ben Ahmed, Benois-Pineau, Allard, Ben Amar, & Catheline, 2014; Chincarini et al., 2011; Chupin et al., 2009; Coupé, Eskildsen, Manjón, Fonov, & Collins, 2012; Gerardin et al., 2009; S. Li et al., 2007; Westman et al., 2011). Despite the recent improvements in detection of AD, the prediction of disease progression using structural MRI alone remains challenging and requires more investigation.

1.5 Problem definition

High-dimensional classification method with higher performance is essential for the success of many applications, especially in automatic classification of patients who suffer from AD. Various high-dimensional pattern recognition algorithms have been introduced a number of neuroimaging studies (I. Beheshti & Demirel, 2015b; Fan, Batmanghelich, Clark, & Davatzikos, 2008; Fan, Shen, & Davatzikos, 2005; Lao et al., 2004). One major issue of high-dimensional classification is the feature-selection method from high-dimensional data to reduce the computational cost and improving the performance. This process is very effective on the final results. In light of this scope, three novel and effective feature selection approaches are introduced in the current thesis to overcome the problem of high-dimensional pattern classification in AD detection.

1.6 Thesis objectives

In this thesis, we propose to use the sMRI data for AD detection. In this context the main objectives are:

- Using voxel-based morphometric (VBM) technique with 3D T1-weighted MRI. The significant local differences of gray matter volume (gray matter atrophies) revealed by VBM analysis are selected as volumes of interests (VOIs). The voxel clusters detected by VBM are employed as VOIs, where each voxel is considered as a feature. This process aids to extracting efficient features in AD detection.
- Use Probability distribution function as novel feature selection method in AD classification. The PDF of a raw feature vector extracted from VOI is a statistical description of the distribution of occurrence probabilities of voxel

values that can be considered a feature vector representing a high-dimensional vector in a lower-dimensional space. Furthermore, we introduce an automatic approach based on the Fisher criterion to determine the optimal number of bins of the histogram generating the PDF.

- Use feature ranking methods as novel feature selection method in high-dimensional AD classification. In this regard, we propose an automatic approach based feature ranking to select discriminative features. In this regard, seven feature-ranking methods, namely, statistical dependency (SD), mutual information (MI), information gain (IG), Pearson's correlation coefficient (PCC), t-test score (TS), Fisher's criterion (FC), and the Gini index (GI are evaluated in proposed feature selection method. It is critical to determine the number of top features. In order to determine the optimal subset features, FC and classification errors are introduced as stopping criteria.
- Compare the generated results with the alternative results of the other methods available in the literature.

1.7 Thesis contributions

High-dimensional classification methods have been a major target of machine learning for the automatic classification of patients who suffer from Alzheimer's disease (AD). One major issue of automatic classification is the feature-selection method from high-dimensional data. In the last decade, several studies investigated high-dimensional pattern classification approach in a number of neuroimaging studies (I. Beheshti & Demirel, 2015b; Fan, Batmanghelich, et al., 2008; Fan et al., 2005; Lao et al., 2004). In the present thesis, we introduce novel feature section

methods in high-dimensional detection of AD. The main contributions of this thesis can be summarized as follows:

1- Utilizing voxel-based morphometric approach, which is one of the best methods for feature extraction from sMRI in AD to detect the MRI voxels that are best, discriminated between the AD group versus HCs (Bron et al., 2015).

2- Introducing a novel statistical feature-selection method based on the probability distribution function (PDF) of the VOI, which can be considered a lower-dimensional feature vector representing sMRI images.

3- Introducing a novel and automatic feature selection method based on feature ranking methods. In this regard, we evaluated seven feature-ranking methods, namely, statistical dependency (SD), mutual information (MI), information gain (IG), Pearson's correlation coefficient (PCC), the *t*-test score (TS), Fisher's criterion (FC), and the Gini index (GI) in the high-dimensional pattern classification. In addition, we introduce three different stopping criteria to determine the optimum number of highest-ranking features (i.e, optimum subset). This procedure helps to determine the relevance of features and class variables and to select the most informative/discriminative features.

4- Introducing data fusion techniques to improve the classification performance, by combining scores or vectors received from clusters obtained from MRI images based on the severity of gray matter atrophy in the brain and during different feature ranking methods.

The experimental results indicate that the performance of the proposed systems are well comparative to that of state-of- the-art classification models.

1.8 Thesis overview

Chapter 2 provides literature review of recent studies in AD detection. Chapter 3 presents the methodology used in this thesis including image acquisition, pre-processing stage and a background of support vector machine as classifier. It also contains the details of the methods with which to assess the classification performance. Chapter 4 describes the probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease. In Chapter 5, we introduce a novel feature selection method based on Feature-ranking and the Fisher criterion to determine the optimal number of top features. In addition, data fusion methods among atrophy clusters are introduced to improve the classification performance. In Chapter 6, we present a novel feature selection method based on Feature-ranking and the classification error to determine the optimal number of top features. The comparison of the proposed methods is provided in Chapter 7. Finally, Chapter 8 presents thesis conclusions on the basis of analysis and discussion and highlights the contributions of this work. It also includes scope for improvement and future direction of research.

Chapter 2

STATE-OF-THE ART IN AD DETECTION

2.1 Introduction

In the last decade, many researchers have investigated to develop automatic computer-aided diagnosis (CAD) system to distinguish AD and HC based on Nuroimaging data. It is worth noting that today's diagnostic procedures are highly dependent on the physician's radiological expertise and are very time-consuming, taking typically a few weeks to complete the evaluation (Petrella, Coleman, & Doraiswamy, 2003). Also, the early diagnosis of AD, which is essential to improve the efficiency of current treatments, is very complex because no characteristic pattern of brain degeneration is well defined, and therefore automated tools may allow a more sensitive analysis and improve diagnostic accuracy. Early detection of AD may help in understanding the root of AD mechanisms as biomarkers for detection and monitoring.

2.2 Biomarkers

Recently, different neuroimaging biomarkers are investigated for AD classification such as X-ray computed tomography (CT), PET, SPECT (Cabral, Morgado, Campos Costa, & Silveira, 2015; Gray et al., 2012; Watanabe, Ono, & Saji, 2015), MRI data (I. Beheshti & Demirel, 2015b; Bron et al., 2015; Kim & Lee, 2013; Savio et al., 2011), Magnetoencephalography (MEG) and Electroencephalography (EEG). Table 2-1 presents a comparison of Nuroimaging techniques in AD classification based on

different biomarkers. Some researchers used unique source of information (I. Beheshti & Demirel, 2015b; Duchesne et al., 2008; Stoeckel et al., 2004; Xia et al., 2008) and some studies combined with each other (Mikhno, Nuevo, Devanand, Parsey, & Laine, 2012; D. Zhang et al., 2011), combined with other clinically relevant data, such as Cognitive Scores, and Mini Mental State Examination (MMSE) (Hinrichs, Singh, Xu, & Johnson, 2011; Westman et al., 2012; D. Zhang et al., 2011; Q. Zhou et al., 2014). This study focuses solely on s-MRI images, because of its noninvasiveness, and its excellent spatial resolution with good tissue contrast, and without radionuclides or radiation exposure, as is observed with PET or SPECT (Beg, Raamana, Barbieri, & Wang, 2012; Matsuda et al., 2012; Nakatsuka et al., 2013).

2.3 Features and feature transformations

Extraction of features from brain images play an important role in the success of classification systems. In general, the type of features can be categorized into two main classes: Using features based on regions of interest (ROI)/volume of interest (VOI) (Gray, Wolz, Keihaninejad, & Heckemann, 2011; Mikhno et al., 2012; Svm, 2008), and using the whole brain ("3D Brain Image-based Diagnosis of Alzheimer's Disease : Bringing Medical Vision into Feature Selection," 2012; Chaves, Ram, Segovia, & Padilla, 2009; Magnin, Mesrob, & Kinkingnéhun, 2009; Moradi et al., 2015; Silveira & Marques, 2010). In the ROI studies, researchers identified the region/volume of brain that are most affected by disease. This approach helps to reduce significantly the dimensionality of feature vectors and select more informative features. Recently, several studies have been using the feature extraction based on ROI/VOI, such as volume of gray matter atrophy (Mikhno et al., 2012; Papakostas et al., 2015; Savio et al., 2011) and shape of hippocampus (Gerardin et

al., 2009; S. Li et al., 2007). Otherwise, using feature extraction based on ROI suffer from defining ROI which is difficult (manual or semi-automatic extraction of regions is unavoidable), time consuming and user dependent task. In contrast, in the whole brain studies, all parts of brain are used in feature extraction procedure, regardless of their meaning that depends on disease. Other feature extraction methods from transformations of the brain volumes, such as Histograms of Gradient Magnitude and Orientation ("Alternative Feature Extraction Methods In 3D Brain Image-Based Diagnosis Of Alzheimer's Disease," 2012), 3D Haar-like features ("Alternative Feature Extraction Methods In 3D Brain Image-Based Diagnosis Of Alzheimer's Disease," 2012), deformation fields (Duchesne et al., 2008) or Normalized Mean Square Error (Chaves, Ram, et al., 2009), are provided in Table 2-2 as literature review. In this thesis, a feature extraction procedure based on VBM analysis is applied to isolate the VOI and Voxel intensity from specific VOIs is used as feature. VBM is an advanced method to assess the whole-brain structure using voxel-by-voxel comparisons (J Ashburner & Friston, 2000; Guo et al., 2010; Matsuda et al., 2012; Moradi et al., 2015; Nakatsuka et al., 2013). It is one of the best methods for feature extraction from sMRI in AD (Bron et al., 2015). More details related to VBM analysis are provided in section 3.3.1.

2.4 Feature selection and dimensionality reduction

Generally, the raw feature space dimensions extracted from nuroimaging data is very high in comparison to the number of samples. Because the sample feature vectors spanned a very small region in the feature vector space, data reduction is desired in post-processing. In this context, it is preferable to reduce the dimensionality of raw feature space. On the other hand, the aim of feature reduction algorithm is to make a set of new features to be used to generate low-dimensional representation of the

original data. In the last decade, many researchers have investigated different dimensionality reduction and feature selection methods such as Principal Component Analysis (PCA) (Illán et al., 2011; Xia et al., 2008), Partial Least Squares (PLS) (Chaves, Ramírez, Górriz, & Puntonet, 2012; Khedher, Ramírez, Górriz, Brahim, & Segovia, 2015; Ramírez et al., 2010; Segovia, Górriz, Ramírez, Salas-González, & Álvarez, 2013) and Linear Discriminant Analysis (LDA) (Ram, Segovia, & Chaves, 2009). In this section, we provide a brief explanation of the mentioned methods.

2.4.1 Dimensionality reduction based on PCA

PCA is a statistical feature dimensionality reduction method. The aim of PCA is to extract a set of orthogonal Principal Components (PCs) from an original data set [26] . Linear combinations of PCs are used to represent high-dimensional original data. Let $X = [X_1, X_2, \dots, X_m]$ where $X_i = (X_{i1}, X_{i2}, \dots, X_{in})^T$ and $i = 1,2, \dots, n$, n is the number of samples. On the other hand, matrix X is defined as follows:

$$X_{n\times m} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & & \vdots \\ x_{n,1} & \cdots & x_{n,m} \end{pmatrix} \tag{2.1}$$

PCs are eigenvectors of the covariance matrix of data X. The covariance matrix is defined as follow:

$$C_{n\times n} = \begin{pmatrix} c_{1,1} & \cdots & c_{1,n} \\ \vdots & \ddots & \vdots \\ c_{n,1} & \cdots & c_{n,n} \end{pmatrix} \tag{2.2}$$

Where $c_{j,k}$ is computed by the following:

$$c_{j,k} = \sum_{i=1}^{m} \frac{(x_{i,j} - \overline{x_j})(x_{i,k} - \overline{x_k})}{m-1} \tag{2.3}$$

Where $\overline{x_j}$ and $\overline{x_k}$ are the average of columns j and k . $\lambda_1 > \lambda_2 > ... > \lambda_n > 0$ are ordered eigen-values of covariance matrix. The eigen-vector (i.e., q) of covariance matrix is defined as follow:

$$Cq = \lambda q \tag{2.4}$$

In the PCA dimensionality reduction, we use the k eigenvectors corresponding to k largest eigenvalues (i.e, $\lambda_1 > \lambda_2 > ... > \lambda_k$), which transfer the dimensionality from n to k as follow:

$$Q = [q_1\ q_2\ \ ...\ \ q_k] \tag{2.5}$$

Where $Q \in R^{m \times k}$.

2.4.2 Dimensionality reduction based on PLS

PLS is a statistical algorithm for modeling the relationship between two datasets: $X \subset R^N$ and $Y \subset R^M$. Recently, the PLS data-reduction approach has been used successfully in a number of applications for machine-learning in AD (Chaves et al., 2012; Khedher et al., 2015; Ramírez et al., 2010; Segovia et al., 2013). After observing n data samples, PLS decomposes the $n \times N$ and the $n \times M$ matrices of zero mean variables X and Y , respectively, into the following form (Segovia et al., 2013; Liang Tang, Peng, Bi, Shan, & Hu, 2014):

$$\begin{aligned} X &= TP^T + E \\ Y &= UQ^T + F \end{aligned} \tag{2.2}$$

where T and U are $n \times A$ matrices of the A extracted score vectors, P and Q are $N \times A$ and $M \times A$ matrices of loadings, and E and F are $n \times N$ and the $n \times M$ error matrices (Segovia et al., 2013). More details about PLS algorithm is provided in section 4.3.3.

Data reduction methods such as PCA and PLS are able to account for combinations of the input features during the process of dimensionality reduction, otherwise in the feature ranking methods only one feature at a time is looked at. But in general, ranking algorithms have lower computational cost compared to data reduction methods. Recently, several studies investigated high-dimensional pattern classification approach in a number of the neuroimaging studies (I. Beheshti & Demirel, 2015b; Fan, Batmanghelich, et al., 2008; Fan et al., 2005; Lao et al., 2004). The contribution of present thesis is to introduce novel feature selection methods for high-dimensional pattern classification in AD.

2.5 Classification methods

Generally, the last stage in AD CAD system is classification and performance evaluation. Recently, several classifiers are introduced in AD classifications such as SVM, neural network (Savio et al., 2011) and Bayesian classifier (Ram et al., 2009; Seixas et al., 2014). In this thesis, we employ SVM classifier for distinguishing AD patients from HC based on supervised learning. Supervised classification based on SVM has been widely used in AD classification (I. Beheshti & Demirel, 2015b; Bron et al., 2015; Dukart et al., 2013; Magnin et al., 2009; Ortiz, Górriz, Ramírez, & Martínez-Murcia, 2013; Savio et al., 2011; Stoeckel & Fung, 2005; L. Zhang, Song, Liu, Bu, & Chen, 2013). Generally, reported accuracies based on SVM learning fall between 80% and 95% (Duchesne et al., 2008; Klöppel et al., 2008; Magnin et al., 2009). More details related to SVM classification and performance evaluation are provided in section 3.4.

Table 2.1: Comparison of different Neuroimaging techniques ("Alzheimer's Association | Alzheimer's Disease and Dementia," 2015).

Biomarker	CT	sMRI	fMRI	MEG	EEG	PET	SPECT
Type	Structural	Structural	Functional	Functional	Functional	Functional	Functional
Radioactivity	No	No	No	No	No	Yes	Yes
Radioactive Tracer	No	No	No	No	No	15O,11C,18F,13N, 82Rb, Pib	99mTc-HMPAO, 99mTc-ECD, 133Xe
Spatial resolution	Low	Good	Good	Good	Good	Good	Good
Cost	Low	Low	Medium	Medium	Low	High	Medium
Stimuli based	No	No	Yes	Yes	Yes	Yes	Yes
Measures	Tissue density	Hemoglobin in the blood	Haemodynamic response (Blood oxygen level)	Neuromagnetic field	Neuroelectrical potentials	Haemodynamic response (CBV, glucose Metabolism)	Haemodynamic response (CBF)
Limitations	-Bone artifacts-May increase risk of cancer -Unable to differentiate tissue types accurately -Unable to visualize the posterior fossa clearly -Measures only anatomy	-Artifacts from non-ferromagnetic metallic objects -Measures only anatomy	-Artifacts from non-ferromagnetic metallic objects -Temporal resolution is limited by the reaction of the body - Expensive, space consuming and immobile scanner -Subjects are not allowed to move at all while being scanned	-Can only measure cortical signals and not those deep inside the brain -Overall brain imaging is beyond its reach -Prone to background noise -Has to be housed in a highly magnetically shielded room -Highly immobile	-Can only measure cortical signals and not those deep inside the brain -Overall brain imaging is beyond its reach -Exerts pressure on subject's head and causes headache -Require application of conductive paste to the skin of head -Background noise can cause significant amount of artifacts	Resolution limited by blood flow -Requires separate session for structural MRI -Repeated scanning is not possible due to use of radioactive tracers	Resolution limited by blood flow -Requires separate session for structural MRI -Repeated scanning is not possible due to use of radioactive tracers -Lower spatial and temporal resolution

Table 2.2: Review of recent studies in AD classification based on different biomarkers

Author(s)	Biomarker(s)	Feature(s)	Feature selection	Learning Algorithm	AD/HC	ACC(%)	SEN(%)	SPE(%)
Stoeckel et al., 2005 [13]	SPECT	Voxel Intensity	-----	SVM	99/31	86.0	84.4	90.9
Duchesne et al., 2008[21]	MRI	Voxel Intensity Deformation field	PCA	SVM	75/75	92.0	-	-
Gorriz et al., 2008 [29]	SPECT	Voxel Intensity	Sub-sampling	SVM	39/41	88.6	-	-
Vemuri et al., 2008 [26]	MRI APOE	Metadata Voxel Intensity	SVM based Wapper	SVM	190/190	89.0	86.0	92.0
Xia et al., 2008 [14]	FDG-PET	Voxel Intensity	PCA Genetic Optimization	SVM	80/70	90.0	-	-
Lopez et al., 2009 [15]	SPECT	Voxel Intensity	PCA+LDA	Gaussian Naive Bayes	42/18	93.4	94.0	92.7
Illan et al., 2010 [16]	PET APOE	Voxel Intensity	PCA	SVM	95/97	88.2	87.8	88.6
Chaves et al, 2012(Chaves et al., 2012)	SPECT	Voxel based features	PCA& PLS	SVM	56/41	91.75	95.12	89.29
Chaves et al, 2012(Chaves et al., 2012)	PET	Voxel based features	PCA& PLS	SVM	75/75	90.00	90.67	89.33
Papakostas et al,2015(Papakostas et al., 2015)	MRI	Voxel Intensity		SVM	49/19	84	90	77
Savio et al, 2011(Savio et al., 2011)	MRI	Voxel Intensity	--	SVM & ANN	49/49	86	80	92

Chapter 3

METHODOLOGY

3.1 Introduction

In this section, a methodology is presented to design an automatic CAD system for MRI classification. This methodology includes image acquisition, preprocessing, classification and performance measurement.

3.2 Image acquisition

MRI images and data used in this work are obtained from the MRI protocol of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [1] . Briefly, the protocol included a 3 Tesla, T1-weighted scanner (Siemens) with Acquisition Plane=SAGITTAL, Acquisition Type=3D, Coil= Phased Arrays (PA), Flip Angle=9.0 degree, Matrix X/Y/Z=240.0 pixels /256 pixels /176 pixels, Mfg Model=Skyra, Pixel Spacing X/Y=1.0 mm/1.0 mm, Pulse Sequence= Gradient Recalled (GR)/Inversion Recovery (IR), Slice Thickness=1.2 mm, and Echo Time (TE) / Inversion Time (TI)/ Repetition Time (TR)=2.98 ms/900 ms/2300 ms.

3.3 Pre-processing

Data pre-processing is the main step in neuroimaging machine learning in order to obtain meaningful results. In this thesis we have used voxel-based morphometry technique in the pre-processing phase. Recently, several studies have been used VBM method for early detection of atrophic changes in AD (I. Beheshti & Demirel,

[1] www.loni.ucla.edu/ADNI

2015a, 2015b; Matsuda et al., 2012; Moradi et al., 2015; Savio et al., 2011) and is introduced as the top feature from sMRI in AD (Bron et al., 2015). In this thesis, data pre-processing is performed using Statistical Parameter Mapping (SPM) software version 8 (Welcome Trust Centre for Neuroimaging, London, UK[1]) and the voxel-based morphometry toolbox version 8 (VBM8[2]), implemented in MATLAB R2014a.

3.3.1 Voxel-Based Morphometry

Morphometry is the technique for investigating of the size, shape and structure of the brain, which is one of the most studied techniques in Neuroimaging. Among the several Morphometry techniques used in brain imaging, such as Voxel-based morphometry (VBM), surface-based morphometry (SBA), deformation-based morphometry (DBM) and tensor based morphometry (TBM). VBM is more widely used in early detection atrophic changes in patients who suffer from AD and is one of the best methods for feature extraction from sMRI in AD (Bron et al., 2015). VBM, introduced by Ashburner and Friston (J Ashburner & Friston, 2000), is a method used to assess whole-brain structure with voxel-by-voxel comparisons, which has been developed to analyze tissue concentrations or volumes between subject groups to distinguish degenerative diseases with dementia (J Ashburner & Friston, 2000; Nakatsuka et al., 2013). Recently, VBM has been applied to detect early atrophic changes in AD (I. Beheshti & Demirel, 2015a; Iman Beheshti, Demirel, & Yang, 2015; Chételat et al., 2005; Hirata et al., 2005; Karas et al., 2003; Matsuda et al., 2012). It can provide statistical results in comparisons of patients with AD to HCs (Baron et al., 2001; Matsuda et al., 2012). Figure 3.1 illustrates overview of VBM on GM component.

[1] http://www.fil.ion.ucl.ac.uk/spm
[2] http://dbm.neuro.uni-jena.de/vbm

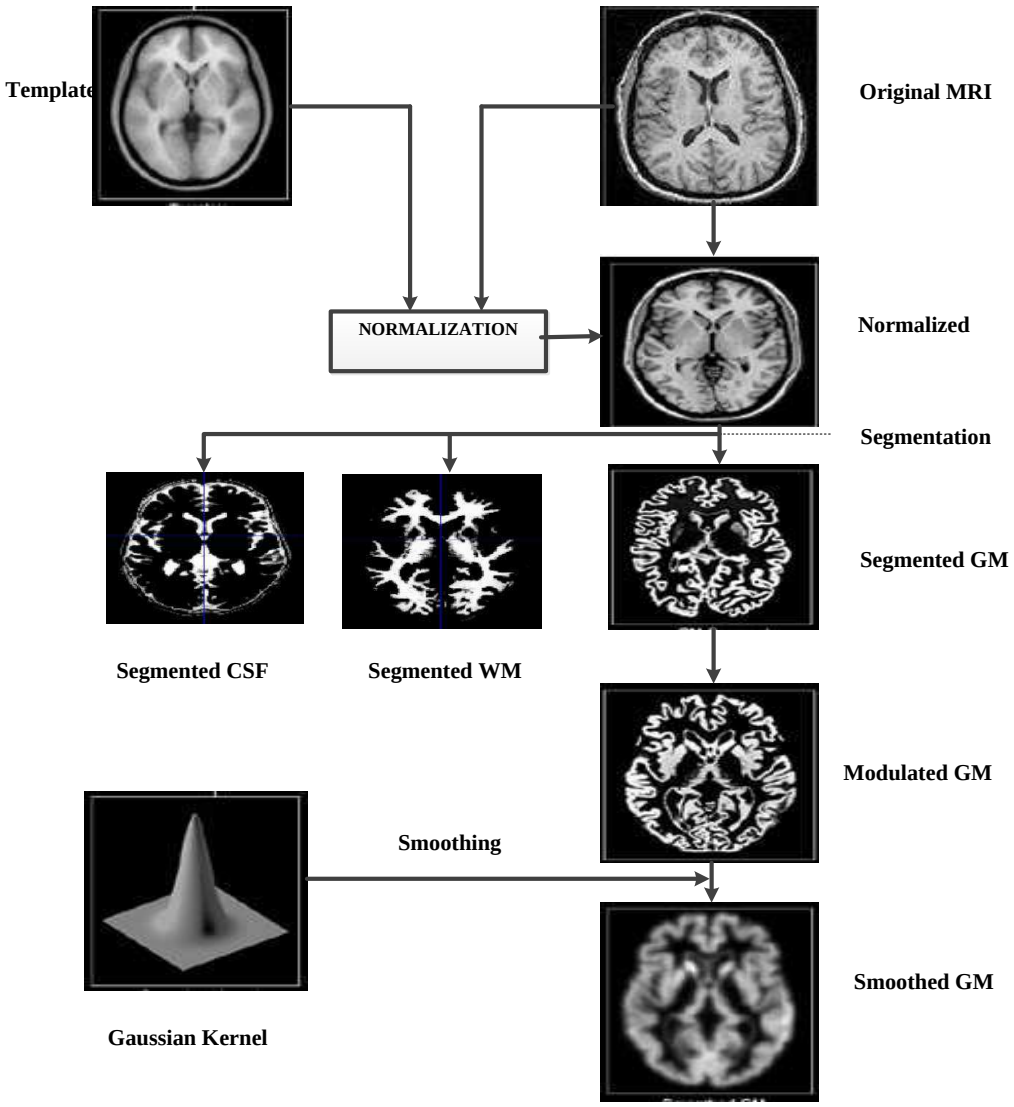


Figure 3.1: The VBM overview processing on GM component

The main steps in VBM processing are as follows:

1- Spatial Normalization: the aim of Spatial Normalization is to provide alignment of MRI images into a standard space (template) in order to establish voxel to voxel correspondence across subjects (Greve, 2011). Figure 3.2: shows the details of Spatial Normalization.

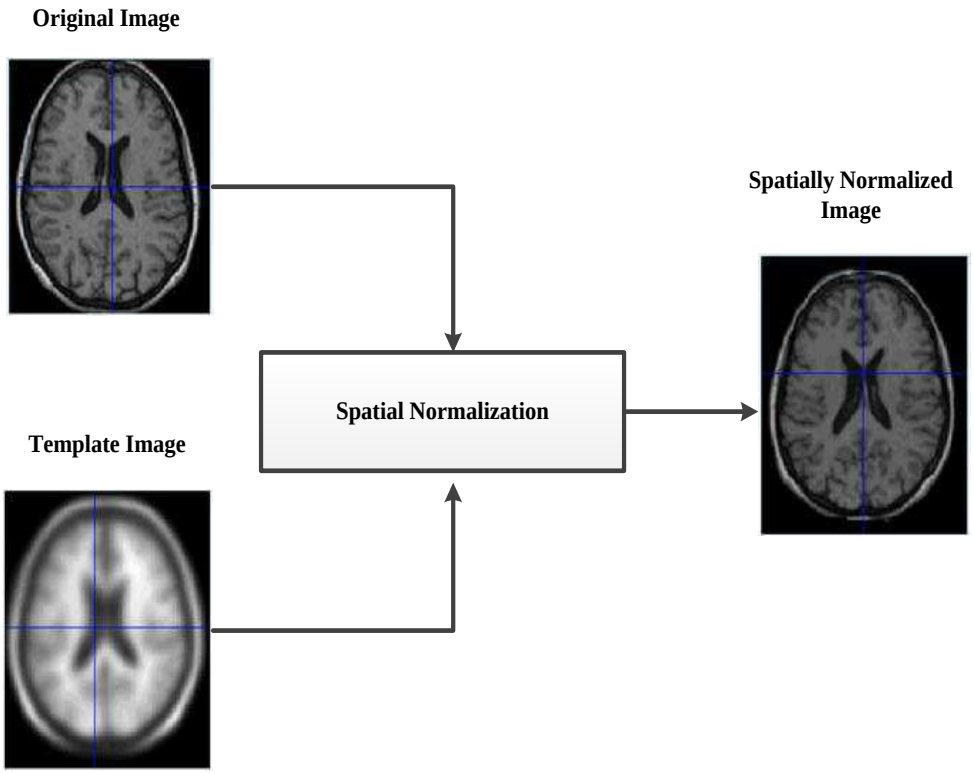


Figure 3.2: The details of Spatial Normalization on MRI. The original MRI is normalized using the template

Segmentation: The aim of segmentation is to segment normalized MRI images into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) components. Figure 3.3 illustrates the segmentation process.

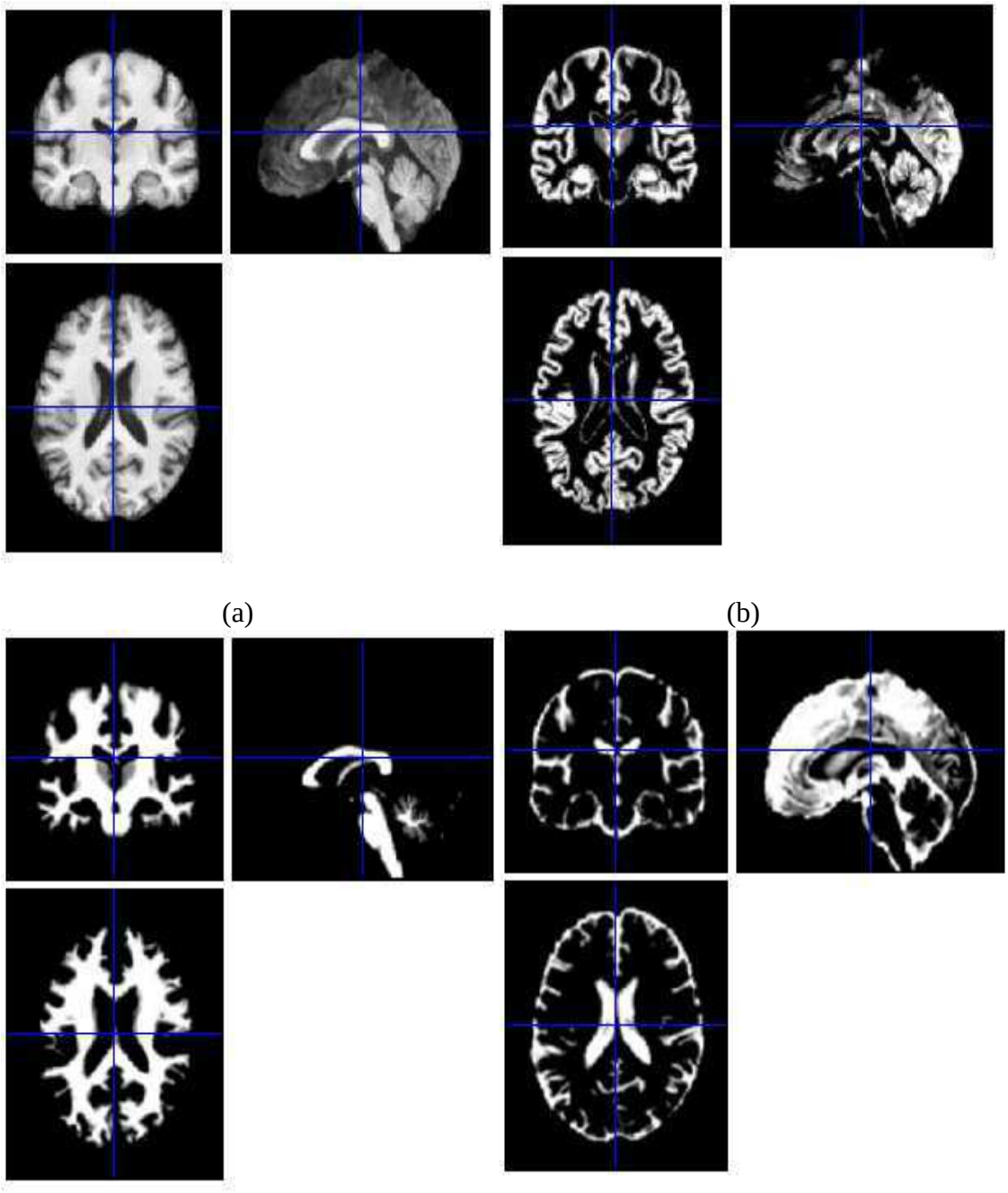


Figure 3.3: The details of segmentation process. (a) Original MRI, (b) segmented GM, (c) segmented WM and (d) segmented CSF

2- Modulation: Modulation step in VBM processing helps to adjust for volume changes during normalization.

3- Smoothing: In spatial smoothing, data points are averaged with their neighborhoods. In this regard, a low pass filter is applied to remove high

frequency components of data while enhancing low frequency components. On the other hand, the aim of smoothing is to increase signal to noise ratio (increasing sensitivity) to prepare images for further processing. In the VBM process, the full-width-half-maximum (FWHM) Gaussian kernel is convolved for spatial smoothing of the MR images. Generally, Gaussian kernel with 6-12 mm FWHM is used for MRI smoothing. Figure 3.4 shows the smoothing process on MRI data.

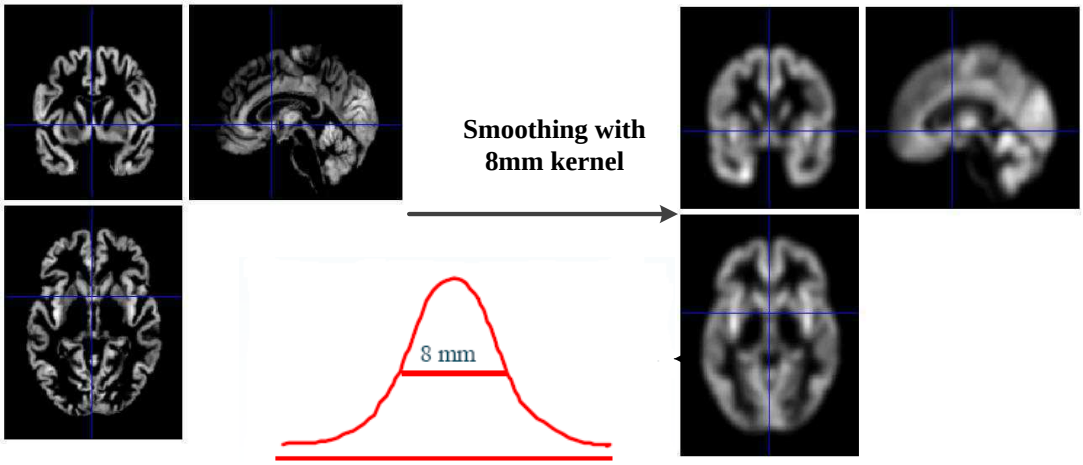


Figure 3.4: The smoothing process on MRI data with Gaussian kernel

In this thesis, we use VBM8 toolbox for voxel-based morphometry processing.

In the VBM8 toolbox, registration to standard Montreal Neurological Institute (MNI) space is an important process, which contains linear affine transformation and nonlinear deformation by using high-dimensional DARTEL normalization. This process involves using the DARTEL template generated from 550 healthy control participants (defined by default settings of VBM8) (Cousijn et al., 2012). Moreover, the DARTEL algorithm provides precise and accurate localization of structural

damage on the MRI images (Matsuda et al., 2012; Nakatsuka et al., 2013). The normalized segmented images are modulated by using a nonlinear deformation, which allows for comparing absolute amounts of tissue corrected for individual differences in brain size (Cousijn et al., 2012). Finally, the segmented images are spatially smoothed with an 8 mm full-width-half-maximum (FWHM) Gaussian kernel. After spatial pre-processing, the smoothed, modulated, DARTEL warped and normalized gray matter datasets are used for statistical analysis. Regional gray matter volume changes are generated by voxel-based analysis over the whole brain. Figure 3.5 illustrates the processing pipeline of the VBM analysis. To detect gray matter volume reductions in patients with AD, a two-sample t-test in SPM8 is used.

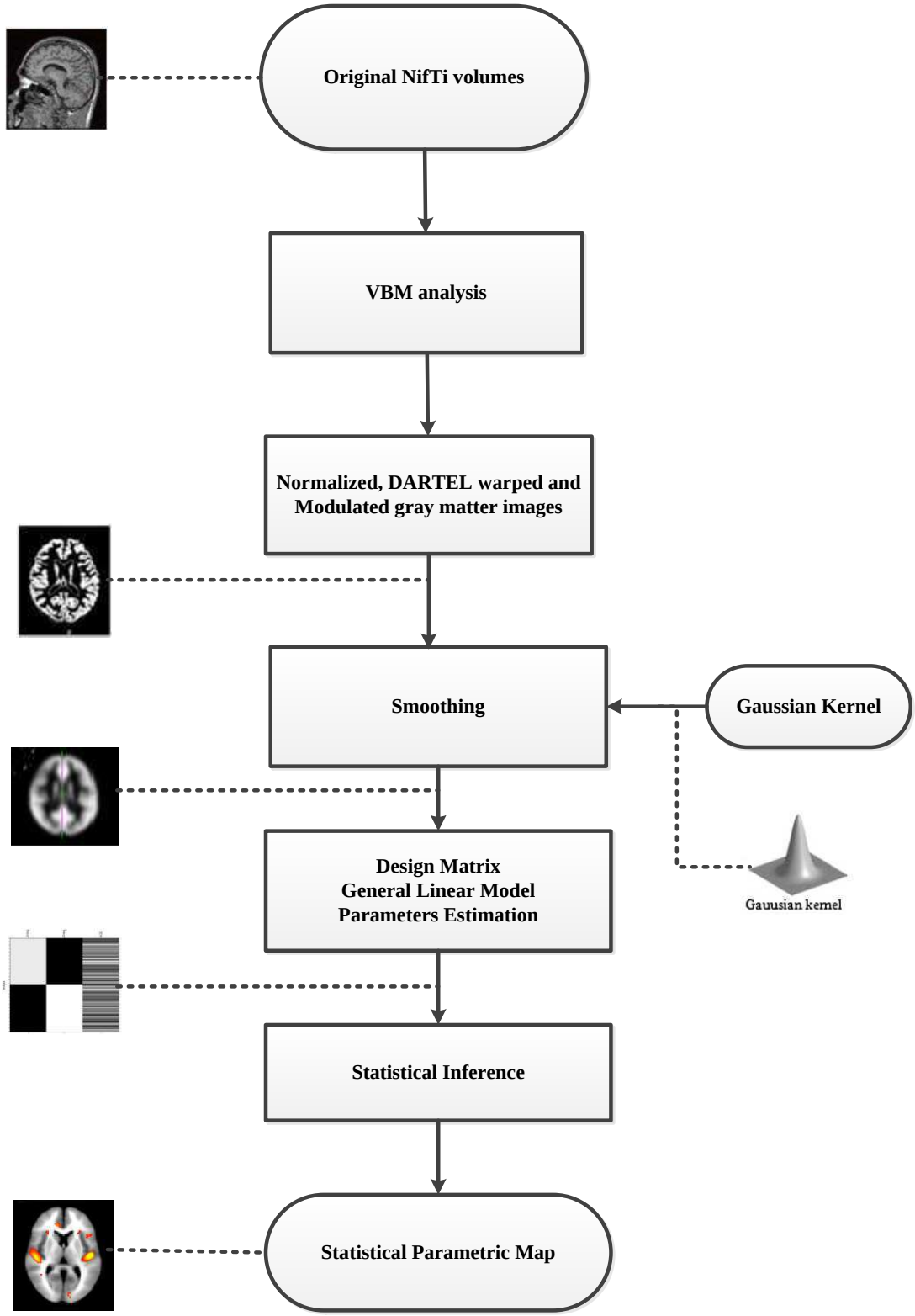


Figure 3.5: The VBM processing pipeline on sMRI data in the present study

Statistical Parameter Mapping (SPM) is an advanced technique to investigate Neuroimaging data such as sMRI, fMRI and PET. In this thesis, we use SPM

software version 8[1] as part of pre-processing in order to investigate the group-wise comparisons between a cross-sectional structural MRI scans diseased group and normal controls. Generally, SPM toolbox uses matrix methods (General Linear Model) relevant to statistical inference (Friston, 2006). A General Linear Model (GLM), can be explained as a variable Y_j based on a linear combinations of the variables as follow:

$$Y_j = x_{j1}\beta_1 + ... + x_{jl} + ... + x_{jL}\beta_L + \varepsilon_j \tag{3.1}$$

where Y_j ($j = 1,...,\mathrm{J}$) is signal intensity at a voxel (as random variable), j is number of observation, x_{jl} ($l = 1,...,L$) is explanatory variable, L is the number of variables, β_l is the unknown parameter corresponding to each x_{jl} and ε_j is noise. In SPM, the two-sample t-test is a special case of GLM, where $Y_{jq} \sim \mathrm{N}(\mu_\mathrm{q}, \sigma_q^2)$ for $q = 1, 2$ are two independent groups of random variables. μ_q and σ_q are the mean and standard deviation of the samples. The GLM can be expressed by matrix notation. By considering equation (3.1) for all observations, we can express:

$$\begin{aligned}
&Y_1 = x_{11}\beta_1 + ... + x_{1l}\beta_l + ... + x_{1L}\beta_L + \varepsilon_1 \\
&\vdots \\
&Y_j = x_{j1}\beta_1 + ... + x_{jl}\beta_l + ... + x_{jL}\beta_L + \varepsilon_j \\
&\vdots \\
&Y_J = x_{J1}\beta_1 + ... + x_{Jl}\beta_l + ... + x_{JL}\beta_1 + \varepsilon_J
\end{aligned} \tag{3.2}$$

which has en equivalent matrix form:

[1] Welcome Trust Centre for Neuroimaging, London, UK; available at: ttp://www.fil.ion.ucl.ac.uk/spm

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_j \\ \vdots \\ Y_J \end{pmatrix} = \begin{bmatrix} x_{11} \cdots x_{il} \cdots x_{1L} \\ \vdots \quad \vdots \quad \vdots \\ x_{j1} \cdots x_{jl} \cdots x_{jL} \\ \vdots \quad \vdots \quad \vdots \\ x_{J1} \cdots x_{Jl} \cdots X_{JL} \end{bmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_J \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_j \\ \vdots \\ \varepsilon_J \end{pmatrix} \tag{3.3}$$

The equation (3.3) can be written in the following form:

$$Y = X\beta + \varepsilon \tag{3.4}$$

Where Y is column vector of observations, β is a column vector on unknown parameters for each voxel ($\beta = [\beta_1, ..., \beta_l, ..., \beta_L]^T$) and ε is the column vector of error terms. The matrix $X, (\mathrm{X} \in R^{J \times L})$ is the matrix design which contains variables indication to which group each image belongs. Figure 3.6 shows an example of the design matrix of the SPM analysis procedure for investigating the differences between the two groups. In the matrix design, each row is one observation and each column is a model parameter. The parameters β are estimated, given $\hat{\beta}$ as follow (Friston, 2006):

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{3.5}$$

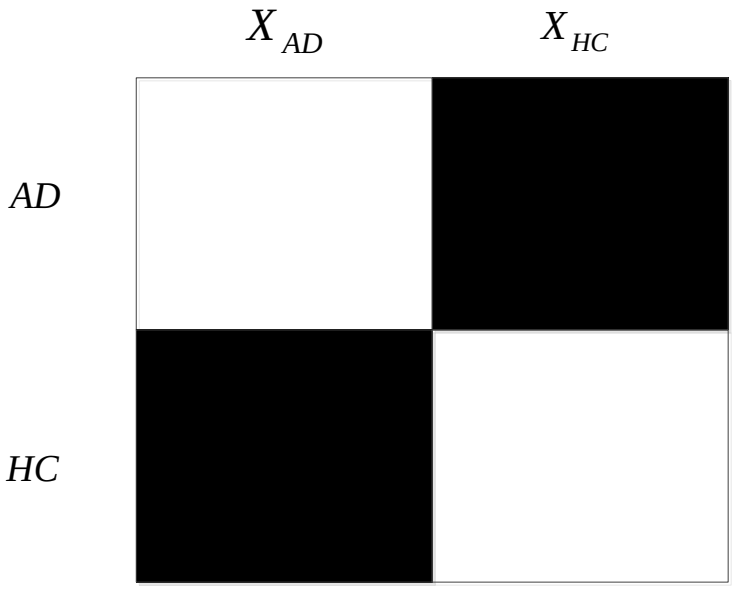


Figure 3.6: an example of the design matrix of the SPM analysis procedure

In the SPM, *t* or *F* statistics between groups are constructed based on linear combination of the parameters $\hat{\beta}$ (contrasts). For example, in the binary case (AD vs. HC), a t-contrast of $[1 \ -1]$ is used to investigate the differential regional effect of AD compared to HC. On the other hand, In order to indentify global and local differences of gray matter in patients with AD compared to healthy controls (HCs), voxel-wise t-statistics is used as follow (Friston, 2006):

$$t = \frac{\beta_{AD} - \beta_{HC}}{SE} \tag{3.6}$$

Where SE is standard error obtained from ε. The equation (3.6) in SPM is equivalent to statistical t-test. Figure 3.7 shows an example of significant voxel differences between patients with AD and age matched HCs. Six 3-D voxel clusters of group comparison representing relative gray matter changes in patients with AD and HCs is shown in Figure 3.8.

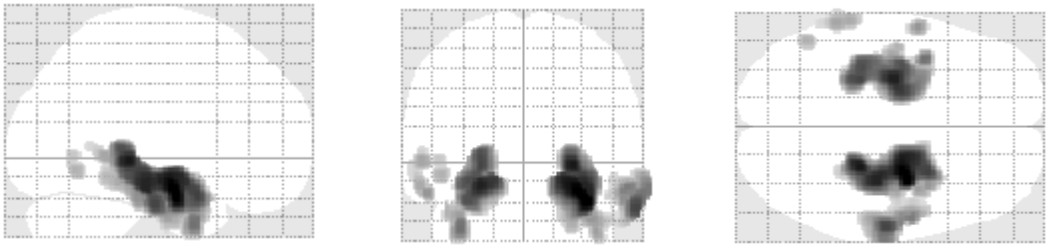


Figure 3.7: Brain regions where there are significant gray matter reduction (atrophy) in patients with AD and age matched HC subjects

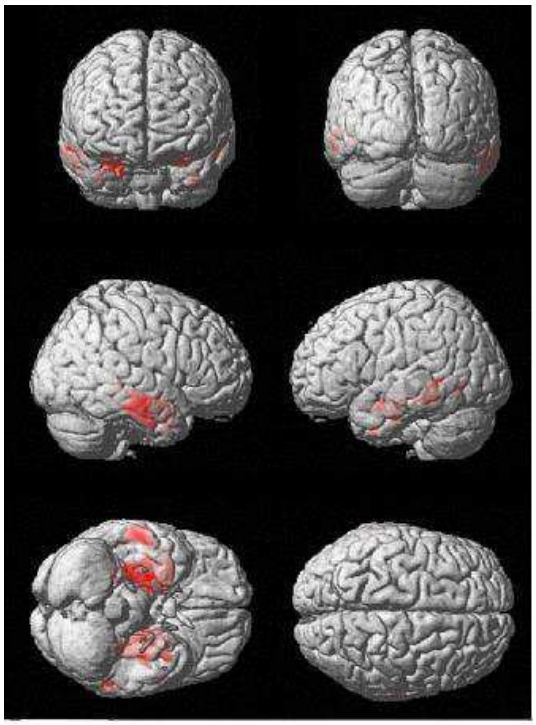


Figure 3.8: Three-dimensional reconstruction of the brain showing gray matter changes in patients with AD and age matched HC subjects. The red region represents the region of gray matter loss

3.4 Classification and performance evaluation

3.4.1 SVM classifier

In this thesis, we classify AD patients apart from HCs by establishing the classification model using the SVM algorithm. The SVM is powerful classifier based on the statistical learning principles. The SVM algorithm has been used successfully in a number of recent applications in machine learning studies (Al-Kadi, 2014;

Dimitrovski, Kocev, Kitanovski, Loskovska, & Džeroski, 2015; Hinrichs et al., 2011; M. Li et al., 2014; Song & Chen, 2014; Xue et al., 2011). During the training, SVM seeks the optimal class-separating hyper-plane in the maximal margin which is the distance between the nearest points (support vectors) on the boundary. Figure 3.9 illustrates of the construction of the SVM hyper plane.

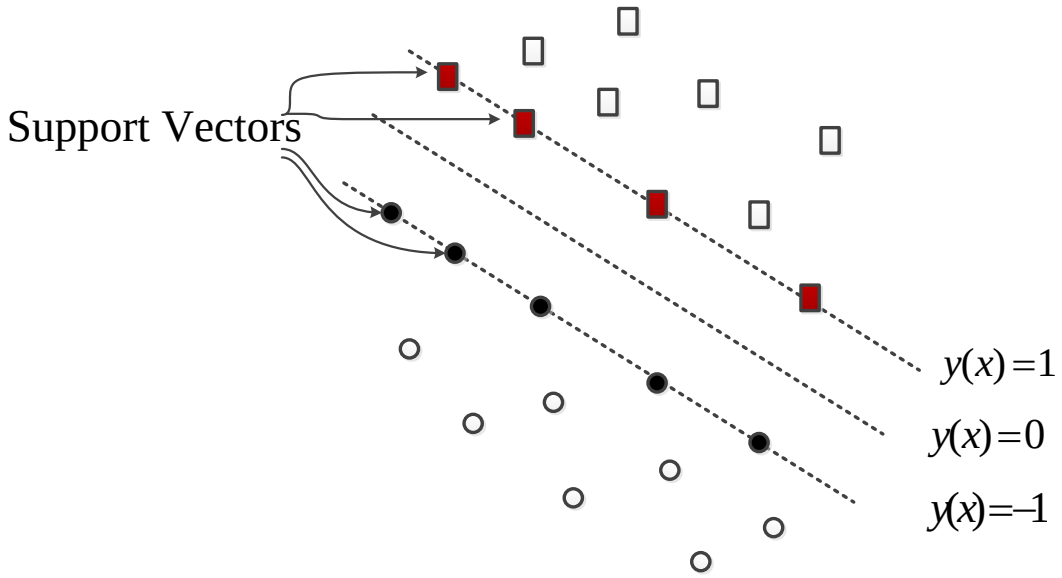


Figure 3.9: Illustration of the construction of the SVM hyper plane

Consider a labeled feature vector, $D=\{\mathrm{X},\mathrm{Y}\}$, where $X \in \Re^p$ (p is the dimension of the input vector) and Y is the class label, which in binary classification with two classes $Y \in \{-1,1\}$. In the SVM classifier, the decision surface is defined as follows:

$$f(x) = sign(\sum \alpha_i y_i K(s_i x) + b) \tag{3.7}$$

where α_i is weight constant, $K(.,.)$ is kernel function, s_i are support vectors and b is bias.

As shown in Figure 3.9, the support vectors are located on the two parallel hyperplanes ($y(\mathrm{x}) = 1$ and $\mathrm{y}(\mathrm{x}) = -1$), where the distance between them is $\frac{2}{\|w\|}$. The maximum distance between the two lines is described as the constrained optimization as follows:

$$\begin{aligned} &\min_{w,b,\xi} \ \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \\ &subject\ to\ y_i (w^T \varphi(x_i) - b) \geq 1 - \xi_i \\ &\xi_i \geq 0, i = 1, 2, ..., n \end{aligned} \tag{3.8}$$

where ξ_i is stack variable. The dual optimization problem is defined as follow:

$$\begin{aligned} &\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\ &subject\ to\ y^T \alpha = 0 \\ &0 \leq \alpha_i \leq C, \mathrm{i} = 1, 2, ..., l \end{aligned} \tag{3.9}$$

Where e is the vector of all ones, l is the number of samples, $C > 0$ is the regularization parameter that needs to be tuned during training and Q is the positive semi-definite matrix with size $l \times l$ as follows:

$$Q_{ij} = y_i y_j K(x_i, x_j) \tag{3.10}$$

where $K(x_i, x_j) = \varphi(\mathrm{x_i})^T \varphi(x_j)$ describes the behavior of support vectors as kernel function. Various kernels can be used during SVM training, such as linear, quadratic, polynomial, and radial basis function (RBF). In this thesis, we use SVM classifier with linear and RBF kernels. Linear and RBF kernel functions are defined as follows, respectively:

$$K(x_i, x_j) = 1 + x_i^T x_j \tag{3.11}$$

$$K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\gamma^2}) \tag{3.12}$$

where, γ is used to controls the kernel width. In this thesis, SVM is performed using LIBSVM[1] and the linear and nonlinear (RBF) kernels.

3.4.2 Validation process

A reliable measurement is achieved by obtaining all the results using the 10-fold cross validation illustrated in Figure 3.10. The RBF model has two parameters that need to be selected: C (regularization) and γ (controls the kernel width); the performance of the classifier depends on these parameters. The C and γ parameters are tuned using the training set, where two cross validation (CV) procedures with grid search are combined. This approach is performed to avoid unwarp bias in the estimation of accuracies produced by the CV procedure (Casanova, Maldjian, & Espeland, 2011). This procedure includes two nested loops. In the outer loop, the data set is split into K_1 folds (K_1=10) at each step: one fold is used as a test and remaining K_1-1 folds for training and validation. In the inner loop, training data (K_1-1 folds) are further divided into K_2 folds (K_2=10). For each combination of C and γ, the classifier is trained using training data and its performance is assessed using the fold remaining for validation by estimating the classification accuracy. One fold is left for validation and the remaining K_2-1 folds are used for training, combined with grid search to determine the optimal parameters. In the grid search, the value of C and γ are varied among the candidate sets $\{2^{-5}, 2^{-4}, ..., 0, ..., 2^{19}, 2^{20}\}$ and $\{2^{-15}, 2^{-14}, ..., 0, ..., 2^{14}, 2^{15}\}$, respectively. The inner loop is repeated K_2 times, measuring the accuracy of the classifier across the K_2 folds for every combination of C and γ. The optimal parameters that produce maximum average accuracy across

[1] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

the K_2 folds are selected, and then the class label of the test data is predicted, which is left out in the outer loop using the selected optimal parameters. The above procedure is repeated K_1 times by leaving a different fold as test data which are used to compute the classification accuracy. For SVM with a linear kernel, only the C parameter is tuned. Over-fitting is prevented by splitting the data into 10 parts, where the training set gets 9 parts and the test set gets 1 part. The data in the training set are used for parameter estimation, whereas the data in the test set are used to measure the performance. This process is repeated 10 times in the context of 10-fold cross validation, where no overlap of the testing sets occurs in this process (Heijden & Ridder, 2004).

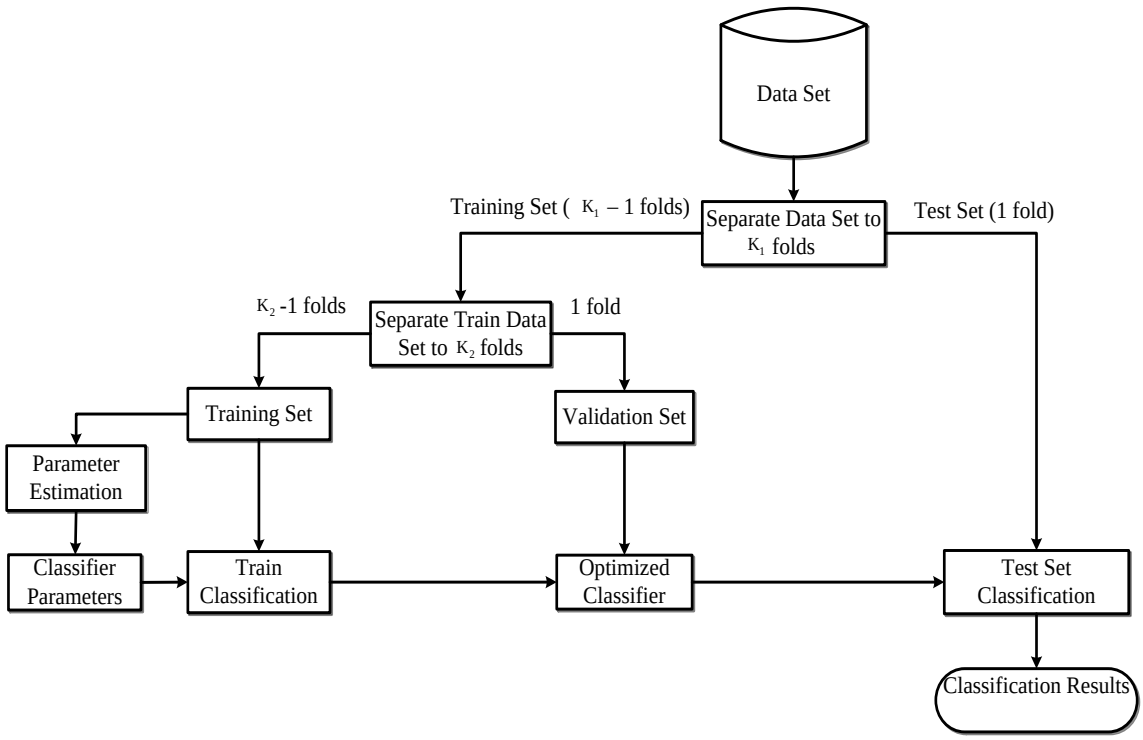


Figure 3.10: 10-fold cross validation method used for parameter tuning and performance testing

3.4.3 Performance evaluation

The classification results are evaluated by means of accuracy (ACC), sensitivity (SEN), specificity (SPE) and area under the curve (AUC), based on 10-fold cross validation. These parameters are defined as follows:

$$ACC = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad 3.13)$$

$$SEN = \frac{TP}{TP + FN} \quad (3.14)$$

$$SPE = \frac{TN}{TN + FP} \quad (3.15)$$

where *TP*, *TN*, *FN*, and *FP* are the number of true positives, true negatives, false negatives, and false positives, respectively. *TP*, *TN*, *FN*, and *FP* are determined as follows:

TP: By counting the number of patients with AD correctly identified as AD.

TN: By counting the number of HCs correctly identified as HCs.

FN: By counting the number of patients with AD incorrectly identified as HCs.

FP: By counting the number of HCs incorrectly identified as AD.

Chapter 4

PROBABILITY DISTRIBUTION FUNCTION-BASED CLASSIFICATION OF ALZHEIMER'S DISEASE

4.1 Introduction

In this chapter, we introduce a novel statistical feature selection method based on the probability distribution function (PDF) of the VOI, which can be considered a lower-dimensional feature vector representing sMRI images. The PDF is assumed to represent the statistical pattern of the VOI representing the entire sMRI. The dimensionality of the PDF-based feature vector can be adjusted by changing the number of bins of the PDF. The proposed PDF-based method not only extracts the selected statistical features but also reduces the dimensionality of the input vectors to feature vectors. The PDF-based feature vector calculation process does not require matrix operations, making the feature extraction process computationally cheaper compared to alternative dimensionality reduction methods such as partial least squares (PLS). In this context, it is apparent that the computational cost of PDF calculation is negligibly low when compared to PLS. The proposed work is accomplished using four steps to develop an automatic computer-aided diagnosis (CAD) technique for AD diagnosis. First, a statistical method is used based on the VBM technique plus Diffeomorphic Anatomical Registration using the Exponentiated Lie algebra (DARTEL) approach to analyze group-wise comparisons between a cross-sectional structural MRI scans diseased group and normal controls (J

Ashburner & Friston, 2000; Cabral et al., 2015; Vemuri & Jack, 2010). Based on the VBM plus DARTEL approach, overall and regional structural gray matter alterations are investigated to define regions with a significant decline of gray matter in patients with AD compared to the healthy controls (HCs). Second, these specified areas (gray matter loss in AD patients) are employed as masks with the template and extracted voxel values from the VOI to form the raw feature vectors. These raw feature vectors go through further data reduction or selection processes before being used by the classifier. Third, a novel statistical feature vector generation using probability distribution functions (PDFs) extracted from the respective 3D mask regions of sMRI is used for classification. The PDF approach can help in two ways: 1) dimensionality reduction and 2) compressing the statistical information of the high-dimensional data into a lower-dimensional vector. PDF pattern recognition has been used successfully in a number of applications, including face recognition (H Demirel & Anbarjafari, n.d.; Hasan Demirel & Anbarjafari, 2008, 2009). In addition, an automatic approach based on the Fisher criterion is used to determine the optimal number of bins of the histogram generating the PDF. This approach adaptively determines the number of PDF bins based on the training data in each fold instead of using a fixed one. Fourth, the performance of the proposed statistical feature-selection technique is evaluated using SVM classifiers.

4.2 Material

4.2.1 Image acquisition

MRI images and data used in this work are obtained from the 3 T MRI protocol of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). Briefly, the protocol included T1-weighted MRI images based on a scanner by Siemens with acquisition plane=sagittal, acquisition type=3D,

coil=PA, flip angle=9.0 degrees, matrix X/Y/Z=240.0/256/176 pixels, mfg model=Skyra, pixel spacing X/Y=1.0/1.0 mm, pulse sequence=GR/IR, slice thickness=1.2 mm, and TE/TI/TR=2.98/900/2300 ms.

4.2.2 Subjects

The group of patients with AD contains 130 people aged 57 to 91 years (mean 75.88±7.54 years). The Mini Mental State Examination (MMSE) and Clinical Dementia Ratio (CDR) scores ranged from 10 to 28 (mean 22.33±3.27) and 0.5 to 2 (mean 0.80±0.37), respectively. The second group contains 130 HCs aged 56 to 88 years (mean 74.49±6.13 years). The MMSE for this group ranged from 27 to 30 (mean 29.26±0.80) and the CDR is zero. In a direct comparison between the HC and AD groups, there are no significant differences in age or the number of gender subjects.

4.3 Methodology of the CAD system

In this section, the methodology is presented based on the PDF approach to design an automatic CAD system for MRI classification. First, the VBM plus DARTEL approach process is used to perform pre-processing on 3D MRI data. Second, a feature-extraction method is employed based on VBM plus DARTEL analysis. Third, an adaptive PDF-based data-selection method is proposed, as a novel statistical data-selection mechanism representing the statistical pattern of VOI of high-dimensional sMRI data in a low-dimensional space. The dimension of the PDF-based vector depends directly on the number of bins used in the histogram of the VOI, which is then normalized into the PDF. The optimal number of bins is obtained by maximizing the Fisher criterion among the possible number of bins. Finally, to evaluate the proposed technique, classifiers such as SVM are used. Figure 4.1 illustrates the framework of the proposed CAD system.

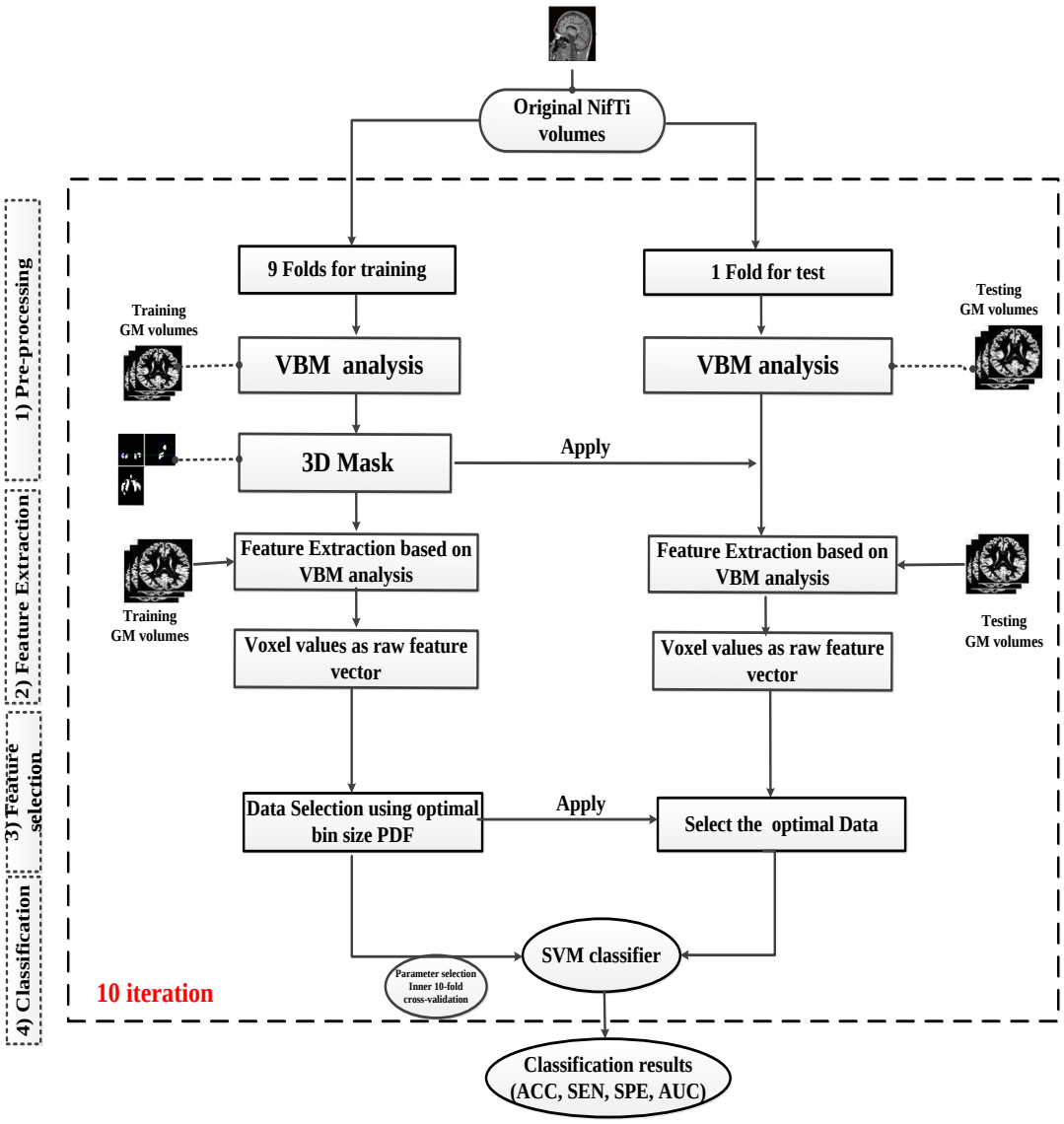


Figure 4.1: The framework of proposed PDF-based CAD system classifying AD

4.3.1 MRI data pre-processing

Data pre-processing is performed using SPM8 and the VBM8 toolbox. VBM, introduced by Ashburner and Friston, is a method used to assess whole-brain structure with voxel-by-voxel comparisons, which has been developed to analyze tissue concentrations or volumes between subject groups to distinguish degenerative diseases with dementia (J Ashburner & Friston, 2000; Nakatsuka et al., 2013). Recently, VBM has been applied to detect early atrophic changes in AD (Chételat et al., 2005; Hirata et al., 2005; Karas et al., 2003; Matsuda et al., 2012). It can provide

statistical results in comparisons of patients with AD to HCs (Baron et al., 2001; Matsuda et al., 2012). To enhance inter-subject registration of the MRI images, DARTEL is applied (John Ashburner, 2007; Modi, Bhattacharya, Singh, Tripathi, & Khushu, 2012), which has been found to optimize the sensitivity of such analyses by using the Levenberg-Marquardt strategy as compared to standard VBM (Klein et al., 2009; Modi et al., 2012). Moreover, the DARTEL algorithm provides precise and accurate localization of structural damage on the MRI images (Matsuda et al., 2012; Nakatsuka et al., 2013). In the VBM8 toolbox, registration to standard Montreal Neurological Institute (MNI) space is an important process, which contains linear affine transformation and nonlinear deformation by using high-dimensional DARTEL normalization. This process involves using the DARTEL template generated from 550 healthy control participants (defined by default settings of VBM8) (Cousijn et al., 2012). The normalized segmented images are modulated by using a nonlinear deformation, which allows for comparing absolute amounts of tissue corrected for individual differences in brain size (Cousijn et al., 2012). Finally, the segmented images are spatially smoothed with an 8 mm full-width-half-maximum (FWHM) Gaussian kernel. After spatial pre-processing, the smoothed, modulated, DARTEL warped and normalized gray matter datasets are used for statistical analysis. Regional gray matter volume changes are generated by voxel-based analysis over the whole brain. To detect gray matter volume reductions in patients with AD, a two-sample t-test in SPM8 is used. Age is applied into the matrix design as a nuisance variable. To avoid possible edge effects between gray matter and white matter or cerebrospinal fluid (CSF), the absolute threshold masking is 0.1. Significance is set at a p-value of <0.01 with correction for family-wise error (FWE) and an extent threshold of 1,400 adjacent voxels for two-sample comparisons.

Between-group differences in demographics and clinical parameters among or between subgroups are executed by Statistical Package for Social Sciences software (SPSS version 16.0) by using an independent sample t-test, and $p<0.05$ is considered significant.

4.3.2 Feature extraction and data reduction and selection

A feature-extraction procedure based on VBM plus DARTEL analysis is applied to isolate the VOI. The regions of decreased gray matter volume obtained using VBM plus DARTEL analysis in patients who suffered from AD are segmented using a 3D mask. This mask is applied to the gray matter density volumes resulting from the VBM plus DARTEL analysis to extract voxel values as raw feature vectors. It is important to separate the data used for VBM 3D mask generation from the data used for classification. In other words, the data to model the 3D mask must explicitly come from the training set. In this context, we divided the dataset for VBM mask generation within each outer cross-validation fold separately. In other words, we randomly divided our subjects into 10 folds with the same number of AD and HC subjects in each fold. In each iteration, we used one fold for testing and 9 folds for training. Based on each training dataset, we performed VBM plus DARTEL analysis to reveal regions of decreased gray matter volume in patients as a 3D mask. In total, we defined 10 different masks with different lengths (e.g. from 59,395 to 69,170 voxels). The respective 3D masks are used in the respective iteration to extract features from the training and testing datasets. The raw feature space in the VBM extracted feature set is very high in comparison to the number of samples. Because the sample feature vectors spanned a very small region in the feature vector space, data reduction is desired in post-processing. In this context, it is preferable to reduce

the dimensionality of sMRI datasets. Therefore, the dimensionality of extracted raw feature vectors is reduced statistically by means of PLS and PDF.

4.3.3 Feature reduction based on PLS

PLS is a statistical algorithm for modeling the relationship between two datasets: $X \subset R^N$ and $Y \subset R^M$. Recently, the PLS data-reduction approach has been used successfully in a number of applications for machine-learning in AD (Chaves et al., 2012; Khedher et al., 2015; Ramírez et al., 2010; Segovia et al., 2013). After observing n data samples, PLS decomposes the $n \times N$ and the $n \times M$ matrices of zero mean variables X and Y, respectively, into the following form (Segovia et al., 2013; Liang Tang et al., 2014):

$$\begin{aligned} X &= TP^T + E \\ Y &= UQ^T + F \end{aligned} \tag{4.1}$$

where T and U are $n \times A$ matrices of the A extracted score vectors, P and Q are $N \times A$ and $M \times A$ matrices of loadings, and E and F are $n \times N$ and the $n \times M$ error matrices (Segovia et al., 2013). In this study, in each fold the PLS algorithm is applied to X (training dataset) and Y (training data label) in order to obtain score and loading matrices. In addition, a weight matrix is obtained from the training dataset to compute a score matrix for the testing dataset (Segovia et al., 2013). Next, score vectors obtained from the training and test datasets are used as feature vectors by SVM classifiers. Figure 4.2 illustrates the pipeline of the PLS feature-reduction procedure.

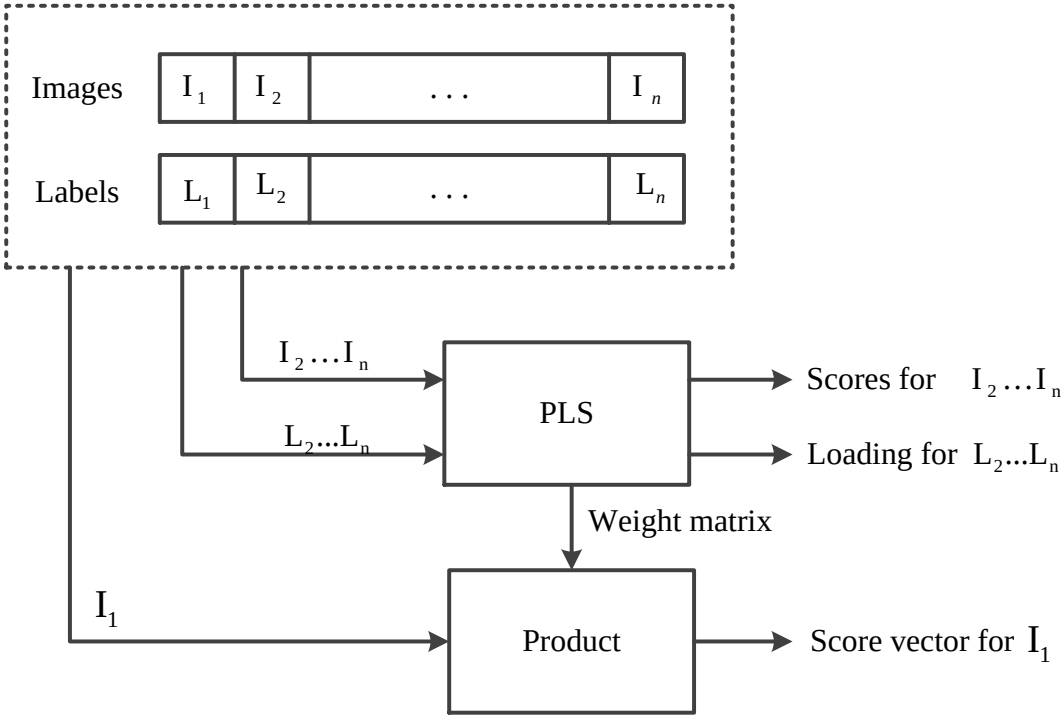


Figure 4.2: Diagram of the PLS based feature extraction (Segovia et al., 2013)

4.3.4 Statistical feature selection based on PDF

The PDF of a raw feature vector extracted from VOI is a statistical description of the distribution of occurrence probabilities of voxel values that can be considered a feature vector representing a high-dimensional vector in a lower-dimensional space. In a mathematical sense, a PDF can be defined as a vector of probabilities representing the probability of the voxel values that fall into various disjointed intervals, known as bins. Given a raw vector extracted from VOI, the PDF, H, of the raw vector met the following conditions (Hasan Demirel & Anbarjafari, 2008, 2009):

$$H = [p_1, p_2, p_3, ..., p_m], \; p_i = \frac{\eta_i}{N}, \; i = 1, 2, ..., m \qquad 4.2)$$

where η_i, is the number of voxels falling into the i^{th} bin, m is the number of bins, and N is the total number of voxels in the 3D mask. In the classification stage, the PDF, H, of raw vectors is used in the representation of the training and test data.

The number of bins adjusts the dimensionality of a PDF vector. In this work, the number of bins is assumed to vary from 2 to 100.

4.3.5 Optimal number of bins based on Fisher criterion

To select the optimal number of bins, an automatic method is used, based on the Fisher criterion, $J(w)$, given in Equation (4.3) :

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{4.3}$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix, respectively (Gao, Liu, Zhang, Hou, & Yang, 2012). For the two classes, C_1 and C_2, the between-class scatter and within-class scatter matrices are defined as:

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \tag{4.4}$$

$$S_W = \sum_{H_i \in C_1} (H_i - \mu_1)(H_i - \mu_1)^T + \sum_{H_i \in C_2} (H_i - \mu_2)(H_i - \mu_2)^T \tag{4.5}$$

where μ_1 is the mean of the PDF vectors in class 1 and μ_2 is the mean of the PDF vectors in class 2, and $w = S_W^{-1}(\mu_1 - \mu_2)$. The main steps in the proposed algorithm are summarized in the pseudo code shown in algorithm 4.1. The number of bins (N_{bin}) of histogram H_i is iteratively incremented from 2 to 100, using a training set of each fold for calculating the respective Fisher criterion values. The optimal number of bins, N_{opt}, maximizing the Fisher criterion is selected to be used as the optimal dimension of the test and training data in each fold through the cross-validation process.

Algorithm 4.1. Optimal number of bins selection procedure

1: $V \leftarrow component\_set(Data_{Train}, Label_{Train})$

2: number of bin$\leftarrow \emptyset$, $N_{bin} = 100$

3: **for** *n = 2 to* N_{bin} **do**

4: $H_i \leftarrow compute\_histogram(X_i, n)$

5: $(S_B, S_W) \leftarrow compute\_scatter(H_i, Label_{Train}))$

6: $\mu_1 \leftarrow mean(H_{i\ class1})$

7: $\mu_2 \leftarrow mean(H_{i\ class2})$

8: $w = S_W^{-1}(\mu_1 - \mu_2)$

9: $\varphi(n) \leftarrow \dfrac{w^T S_B w}{w^T S_W w}$

10: **end for**

11: $N_{opt} \leftarrow \arg\max_{n \in \{2,...,N_{bin}\}} \varphi(n)$

4.4 Experimental results and discussion

In this section, the experimental results of VBM plus DARTEL analysis on 3D MRI are reported to reveal the significance of the volumetric regions with atrophy in patients, contributing to VOI. The performance of the classification of AD using a 10-fold cross-validation is also presented for four cases: 1) performance of the raw features (VBM features) dataset, 2) performance of the PLS method, 3) performance of the proposed PDF technique, and 4) performance of the PDF technique using the optimal number of bins. Two types of SVM classifiers, namely SVM-linear and SVM-RBF, are used for AD classification. *ACC* (%), *SEN* (%), *SPE* (%), and *AUC* (%) performance metrics are used to assess the different scenarios.

4.4.1 Voxel-based morphometry on gray matter

VBM plus DARTEL revealed a significant decline of gray matter volume in the right hippocampus, left hippocampus, right inferior parietal lobe, and right anterior cingulate in patients with AD compared to the HCs. Figure 4.3 shows the brain regions where there is significant atrophy in gray matter volume in AD patients compared to HCs in fold 1 training. The voxel locations of these significant regions are used as a 3D mask in each fold. This mask is applied to the gray matter density volume results from the segmentation step in the VBM plus DARTEL analysis to extract voxel values as raw feature vectors.

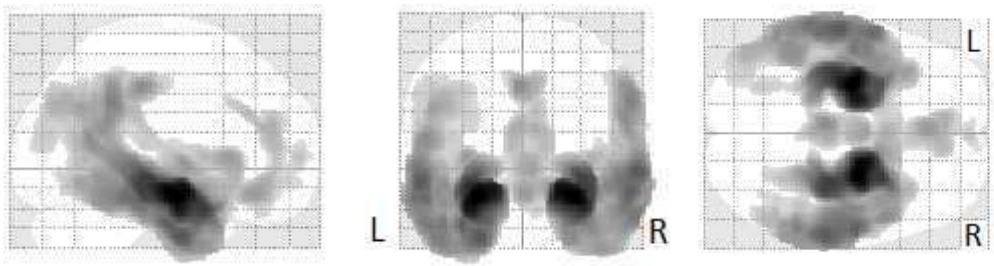


Figure 4.3. Comparison of gray matter volume among 117 patients with AD and 117 HCs in fold 1 training by VBM using SPM8 (FWE corrected at $p < 0.01$ and extend threshold $K = 1400$)

4.4.2 Performance of raw feature representation

The complete MRI dataset from the ADNI database consisted of 260 samples. Table 4.1 presents the ACC, SEN, SPE, and AUC obtained by 10-fold cross validation using SVM-linear and SVM-RBF classifiers for raw feature vectors obtained by masking after VBM plus DARTEL analysis.

Table 4.1: Performance comparison on VBM features data sets on 10 fold cross validation for raw feature vectors

Classifier	ACC(%)	SEN(%)	SPE(%)	AUC (%)
SVM-linear	83.58	82.04	85.12	92.10
SVM-RBF	86.02	89.70	82.35	93.13

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; SVM, Support Vector Machine; RBF, Radial Basis Function.

4.4.3 Performance of PLS method

The feature reduction using PLS is accomplished by extracting raw feature data from VOI obtained from VBM analysis. The extracted raw feature vectors are reduced to lower-dimensional feature vectors of up to 100 components using PLS. Table 4.2 (a) presents the *ACC*, *SEN*, *SPE*, and *AUC* obtained from 10-fold cross-validation for SVM classifiers for changing dimensionality. According to Table 4.2(a), it is clear that the maximum accuracy (90.76%) is yielded with SVM-RBF when the dimensionality is 80. The accuracy is 4.74% higher than the same classifier with all raw features used in Table 4.1. The reset of the results in Table 4.2 (a) are also higher than the raw data for *SEN*, *SPE,* and *AUC*. The results reported in Table 4.1 and Table 4.2 (a) indicate that the PLS performance using SVM-linear and SVM-RBF classifiers is higher than with the raw data.

4.4.4 Performance of proposed PDF-based technique

The feature selection using PDF is accomplished by extracting raw feature data from VOI obtained using VBM analysis. The extracted raw feature vectors are reduced to lower-dimensional feature vectors of up to 100 components by changing the number of bins of the PDF. Table 4.2 (b) and Figure 4.4 present the *ACC*, *SEN*, *SPE*, and *AUC* obtained by 10-fold cross-validation using SVM-linear and SVM-RBF classifiers. The results reported in Table 4.2 (a) and Table 4.2 (b) show that the PDF-

based method is with higher ACC than the PLS-based method in most of the dimensions using linear and SVM-RBF classifiers. For example, for 20 components, the PDF-based ACC performance is 88.50% while PLS ACC performance is 81.96% using SVM-linear. There are few cases in which PLS ACC is higher. The same observation is valid for AUC and SPE, where the PDF-based method is mostly superior to the PLS-based method. On the other hand, although for SEN the PDF-based method is better than the PLS-based method in SVM-linear, the PLS-based method is higher for the SVM-RBF classifier.

Table 4.2: Performance analysis of the PDF based method in comparison to PLS based method

(a) Performance comparison on PLS reduced features data sets on 10 fold cross validation

No. of components	ACC(%)	SEN(%)	SPE(%)	AUC(%)	Classifier
2	**87.34**	**84.65**	**90.03**	**95.33**	
10	85.42	81.57	89.26	93.31	
20	**81.96**	**81.57**	**82.34**	**92.25**	
30	81.19	80.03	82.34	91.66	
40	81.96	80.03	83.88	92.19	SVM
50	82.73	80.03	85.42	92.49	Linear Kernel
60	82.73	80.03	85.42	92.66	
70	83.88	82.34	85.42	92.90	
80	84.26	82.34	86.19	93.14	
90	85.03	83.88	86.19	93.26	
100	85.03	83.88	86.19	93.31	
2	86.53	88.46	84.61	91.60	
10	74.61	96.15	53.07	90.41	
20	79.23	94.61	63.84	93.20	
30	86.76	93.07	78.46	94.50	
40	88.84	92.30	85.38	94.73	SVM
50	88.07	90.76	85.38	95.27	
60	88.46	90.76	86.15	95.38	RBF Kernel
70	90.38	90.76	90.00	95.74	
80	**90.76**	**90.76**	**90.76**	**95.86**	
90	90.76	90.76	90.76	95.92	
100	90.76	90.76	90.76	95.92	

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; SVM, Support Vector Machine; RBF, Radial Basis Function.

Table 4.2: Performance analysis of the PDF based method in comparison to PLS based method

(b) Performance comparison on PDF reduced features data sets on 10 fold cross validation

No. of component	ACC(%)	SEN(%)	SPE(%)	AUC(%)	Classifier
2	86.19	83.88	88.50	94.85	
10	87.73	86.19	89.26	95.62	
20	**88.50**	**86.19**	**90.80**	**95.50**	
30	88.50	88.50	88.50	94.73	
40	86.96	83.88	90.03	94.91	SVM
50	87.34	86.96	87.73	94.91	Linear Kernel
60	88.11	86.96	89.26	95.15	
70	87.34	87.73	86.96	95.21	
80	88.50	89.26	87.73	95.80	
90	86.96	86.96	86.96	94.62	
100	88.50	87.73	89.26	96.21	
2	88.07	87.69	88.46	95.86	
10	89.61	89.23	90.00	96.39	
20	88.84	86.92	90.76	96.51	
30	90.00	90.00	90.00	96.09	
40	88.84	88.46	89.23	95.92	SVM
50	89.61	88.46	90.76	96.15	
60	89.61	90.00	89.23	96.27	RBF Kernel
70	88.84	86.92	90.76	96.21	
80	90.00	90.00	90.00	96.75	
90	90.00	88.46	91.53	96.75	
100	**90.76**	**90**	**91.53**	**97.04**	

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; SVM, Support Vector Machine; RBF, Radial Basis Function.

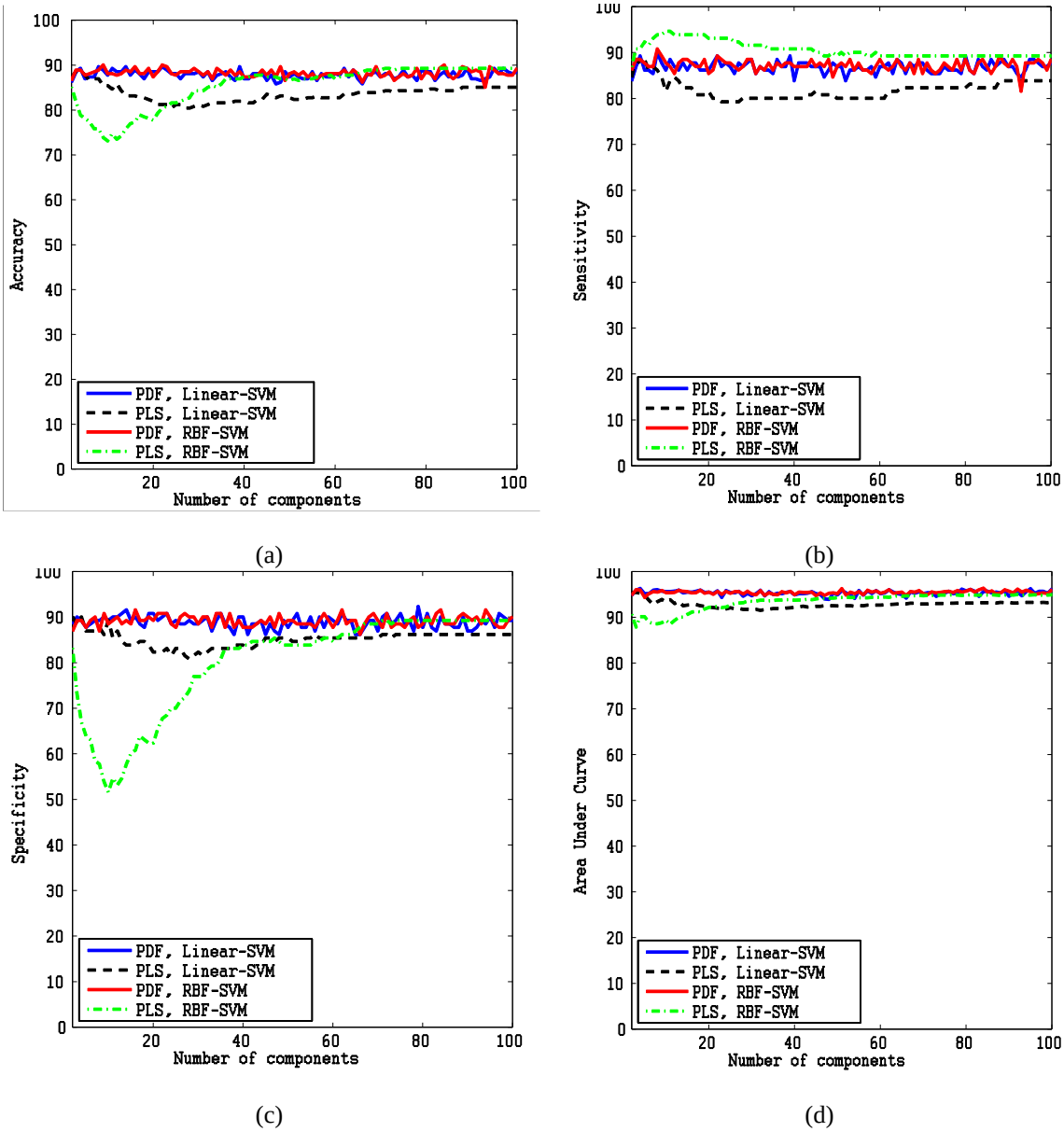


Figure 4.4: Classifier performance based on PLS and PDF feature selection: (a) Accuracy(%), (b) Sensitivity(%), (c) Specificity(%) and (d) Area Under Curve(%)

4.4.5 Performance of PDF technique using optimal number of bins

As proposed in section 4.3.5, the optimal number of bins is determined by maximizing the Fisher criterion applied to the two classes (AD and HC) of the training data in each fold through the cross-validation process. Table 3 presents the average of the performances of the classifiers with the optimal number of bins obtained in each fold, through 10-fold cross-validation. The proposed method of

determining the optimal number of components (i.e. the number of bins) is also applied to PLS. By examining the results of Table 4.3, it is observed that the overall performance of the proposed PDF-based method with the optimal number of bins is superior to PLS for SVM-linear, where the results of both methods are comparable for SVM-RBF.

Table 4.3: Performance results of the PDF and PLS based methods with optimal number of bins

Classifier	ACC(%)	SEN(%)	SPE(%)	AUC (%)
PDF-SVM-linear	**89.65**	**87.73**	**91.57**	**95.33**
PDF-SVM-RBF	88.83	87.73	90.03	95.39
PLS-SVM-linear	85.42	84.65	86.19	93.32
PLS-SVM-RBF	89.26	89.26	89.26	95.09

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; PDF, Probability Distribution Function; PLS, Partial Least Squares; SVM, Support Vector Machine, RBF, Radial Basis Function.

4.5 Performance comparison to other methods

Recently, several studies have reported classification results to distinguish AD and HC based on MRI. Zhang et al. (D. Zhang et al., 2011) used multimodal classification of AD based on the combination of MRI, CSF, and PET. They reported ACCs of 86.2%, 82.1%, and 86.5% in the classification of AD/HC by MRI, CSF, and PET imaging modalities, respectively. Also, they achieved a high accuracy performance (93.2%) by combining the MRI, CSF, and PET results. Querbes et al. (Querbes et al., 2009) achieved an ACC of 85% based on the cortical thickness feature from MRI data. Hinriches et al. (Hinrichs, Singh, Xu, & Johnson, 2009) reported an ACC of 75.27% based on MRI data and increased it to 81% by combining MRI and PET. Vemuri et al. (Vemuri et al., 2008) announced an

SEN/SPE of 86/86% in 380 subjects using the STructural Abnormality iNDex (STAND) score from MRI data. Westman et al. (Westman et al., 2012) presented an ACC of 87% from MRI data and increased it to 91.8% by combining MRI data with CSF measures. Papakostas et al. (Papakostas et al., 2015) applied two methods to analyze MRI data, namely, VBM and deformation-based morphometry (DBM), on 98 female subjects. They extracted features based on three different models: MSD, displacement magnitude (DM), and Jacobian determinant (JD). They also investigated their methods with several classifiers. They reported ACCs of 85%, 84%, and 79% for the three models, respectively. Aguilar et al. (Aguilar et al., 2013) used FreeSurfer software to compute cortical thickness and volumetric measures, yielding an ACC of 84.9% for the artificial neural network (ANN) classifier from MRI data and of 88.8% for the SVM classifier by combining MRI data with educational and demographic data. Zhou et al. (Q. Zhou et al., 2014) employed FreeSurfer software to calculate 55 volumetric variables from MRI data. They reported an ACC of 78% for MRI data and 92.4% by combining MRI data with the MMSE. Savio et al. (Savio et al., 2011) studied the feature-extraction process with VBM analysis on 98% female subjects only and achieved the best results with 86% accuracy for the RBF-AB-SVM classifier. Khedher et al. (Khedher et al., 2015) reported an ACC of 88.49% by combining GM and WM modalities in MRI. Kloppel et al. (Klöppel et al., 2008) employed leave-one-out as a validation method in three different groups (Groups I, II, and III) with different severity of atrophy in AD. The ACC of Group I is 95%, of Group II is 92.9%, and of Group III is 81.1%. The severity of atrophy in Group I is the highest, making this group the most successful among the three. A study by Cuingnet et al. (Cuingnet et al., 2011) comprised 10 methods using the ADNI database. They reported a SEN of 81% and a SPE of 95%

as the best performances. In this chapter, a set of a total of 260 MRI samples is used in the AD and HC groups, with superior results with respect to ACC, SEN, and AUC in Table 4 except for the results of Kloppel et al. (Klöppel et al., 2008) for Groups I and II. One of the main reasons for this observation stems from the fact that the severity of the atrophy of Groups I and II is higher than that of Group III and our dataset. Additionally, using the leave-one-out method already gives an advantage to the method employed by Kloppel et al. (Klöppel et al., 2008) against the 10-fold cross-validation technique used in the proposed method. The experimental results using the proposed PDF-based approach with SVM by linear Kernel generates 89.65% accuracy, 87.73% sensitivity, 91.57% specificity, and 95.33% AUC. The details of the parameters used in classification performance with different methods by using MRI data are provided in Table 4.4. Some of the results reported in Table 4.4 use ADNI data-set, where the others use different or private data-sets. Additionally, the results from ADNI data-set are using different number of AD/HC samples. In order to have comparable results, we have used ADNI data-set with high number of samples (130 AD and 130 HC), which we believe provides a suitable ground for acceptable comparisons.

Table 4.4: Supervised classification results of Alzheimer's disease and healthy control subjects on MRI data

Author	Source of data	AD/HC	Validation method	ACC(%)	SEN(%)	SPE(%)	AUC (%)
Zhang et al.,2011(D. Zhang et al., 2011)	ADNI	51/52	10 Fold	86.20	86.00	86.30	-
Querbes et al.,2009 (Querbes et al., 2009)	ADNI	130/130	10 Fold	85.00	-	-	-
Hinrichs et al., 2009 (Hinrichs et al., 2009)	ADNI	77/82	10 Fold	75.27	63.06	81.86	82.48
Vemuri et al., 2008 (Vemuri et al., 2008)	ADRC /ADPR	190/190	4 Fold	-	86.00	86.00	-
Westman et al., 2012 (Westman et al., 2012)	ADNI	96/111	10 Fold	87.00	83.30	90.10	93.00
Papakostas et al, 2015 (Papakostas et al., 2015)	OASIS	49/49	10 Fold	85.00	78.00	92.00	-
Aguilar et al.2013 (Aguilar et al., 2013)	ADNI	116/110	10 Fold	84.90	80.20	90.00	88.00
Zhou et al.,2014 (Q. Zhou et al., 2014)	ADNI	59/127	2 Fold	78.20	68.50	81.70	-
Savio et al., 2011(Savio et al., 2011)	OASIS	49/49	10 Fold	86.00	80.00	92.00	-
Khedher et al., 2015 (Khedher et al., 2015)	ADNI	188/229	10 Fold	88.49	85.11	91.27	-
Kloppel et al., 2008 (Klöppel et al., 2008)	Private	20/20	Leave-one-out	95.00	95.00	95.00	-
Kloppel et al., 2008 (Klöppel et al., 2008)	Private	14/14	Leave-one-out	92.90	100	85.70	-
Kloppel et al., 2008 (Klöppel et al., 2008)	Private	33/57	Leave-one-out	81.10	60.60	93.00	-
Cuingnet et al., 2011(Cuingnet et al., 2011)*	ADNI	162/137	2 Fold	-	81.00	95.00	-
Proposed method	**ADNI**	**130/130**	**10 Fold**	**89.65**	**87.73**	**91.57**	**95.33**

*This paper by Cuingnet et al. Compares ten methods and the best performance is given here.

4.6 Conclusion

In this chapter, an automatic CAD technique is introduced based on a novel statistical feature-selection process, namely, PDF of VOI, for the classification of AD. The proposed feature-selection method compresses the statistical information of

high-dimensional data into a lower-dimensional vector. This approach is used for high-dimensional classification, especially for feature-extracted VOI of gray matter atrophy. The PDF-based feature-selection approach is compared to the standard PLS-based classification using SVM classifiers. The results clearly indicated that the PDF-based feature-selection method is a reliable alternative to the PLS-based method, in which the performance of the proposed PDF-based method with the optimal number of bins is superior to PLS for SVM-linear, and the results of both methods are comparable for SVM-RBF. Moreover, PDF generation does not require complex matrix operations, making the feature-extraction process computationally cheaper than alternative dimensionality-reduction methods, such as PLS. The proposed PDF-based method not only extracts the selected statistical features but also reduces the dimensionality of the input vectors to feature vectors with acceptably low dimensions. It is apparent that the computational cost of PDF calculation is negligibly low when compared to PLS. As part of future prospects on PDF-based pattern recognition in neuroimaging, it is suggested to use data fusion techniques for the proposed MRI modality with other modalities, such as PET, CSF, and WM, and to combine them using the proposed PDF-based approach in order to achieve higher accuracy. The PDF-based data fusion technique has already been used successfully in recent studies for the improvement of face-recognition performance (H Demirel & Anbarjafari, n.d.; Hasan Demirel & Anbarjafari, 2009).

Chapter 5

STRUCTURAL MRI-BASED DETECTION OF ALZHEIMER'S DISEASE USING FEATURE RANKING AND FISHER CRITERION

5.1 Introduction

This chapter describes the use of a statistical feature ranking approach using t-test as part of a novel feature selection process. The number of highest ranking features selected is determined by using the Fisher Criterion, which maximizes the class separation between AD and HC groups. The Fisher Criterion aids in finding an optimal number of features with the most discriminative information for the classification process. The proposed feature selection method is applied to different atrophy clusters of voxels, which correspond to the volumes of interest (VOIs) in the gray matter of the MRI obtained through the voxel-based morphometry (VBM) analysis in the preprocessing. In this context, data fusion is introduced to increase the classification performance, which utilizes a majority-voting-based score fusion and a feature vector concatenation-based source fusion. In the proposed system, we use only MRI data, unlike several recent studies where MRI is combined with other different data such as PET, Cognitive Scores, and Mini Mental State Examination(MMSE) to increase the classifier performance (Hinrichs et al., 2011; Westman et al., 2012; D. Zhang et al., 2011; Q. Zhou et al., 2014). The proposed system is accomplished by the systematic use of several ideas at five levels. At the

first level, the VBM technique is employed to analyze group-wise comparisons between cross-sectional structural MRI scans, in order to find the MRI voxels that are best discriminated between the AD group and the HC group (J Ashburner & Friston, 2000; Matsuda et al., 2012; Moradi et al., 2015; Nakatsuka et al., 2013). The inter-subject registration of the MRI images is promoted by employing the Diffeomorphic Anatomic Registration Through Exponentiated Lie algebra algorithm (DARTEL) (Matsuda et al., 2012). This algorithm provides precise, accurate localization of structural damage of the MRI images (Matsuda et al., 2012; Nakatsuka et al., 2013). Based on the VBM plus DARTEL approach, the overall and regional structural gray matter alterations are investigated to define regions with significant atrophy of gray matter in the patients who suffer from AD. The results obtained from 68 patients with AD, when compared to 68 HCs, show significant gray matter decline in right/left hippocampuses and in the inferior parietal and anterior cingulate regions in patients with AD. Instead of making a single global classifier, the multiple individual classifiers based on atrophy clusters obtained using VBM plus DARTEL analysis are proposed for use with data fusion techniques for more accurate classification. Based on these clusters, five different VOIs are defined as follows: 1) VOI_1 includes the right hippocampus region, 2) VOI_2 includes the left hippocampus region, 3) VOI_3 contains the right inferior parietal lobule region, 4) VOI_4 includes the right anterior cingulate region, and 5) VOI_{all} contains an accumulation of all atrophy cluster regions. At the second level, specified VOIs are used as 3D masks to extract voxel values from the VOIs to generate raw feature vectors. These raw feature vectors can be used in the data selection processes before use by the classifiers. At the third level, the extracted features are systematically ranked, based on the t-test values of the respective features obtained from the

training set. The t-test can be considered as a statistical indicator showing the level of separation/discrimination between two groups (AD and HC) in the training set. For this reason, ranking according to the t-test, followed by the use of a subset of highest ranking features, would increase the classification performance. The t-test feature ranking has been used successfully in a number of pattern recognitions studies(Chaves, Ramírez, et al., 2009; M. Liu, Zhang, & Shen, 2012; D. Wang, Zhang, Liu, Lv, & Wang, 2014). In addition, an automatic approach based on the Fisher Criterion is proposed to determine the number of top features. This approach adaptively determines the optimum number of top features and identifies a discriminative subset of high performance features based on training data in each fold, instead of using a fixed number of features. At the fourth level, the performance of the proposed feature selection technique is evaluated using support vector machine (SVM) classifiers. In the present work, the SVM classifier with both linear (linear SVM) and nonlinear (RBF SVM) kernels is trained to discriminate between the classes. In the final level, data fusion techniques among atrophy clusters (VOIs) are proposed to increase the overall performance. Data fusion improves the classification performance by integrating data (vectors, classifiers) from different atrophy clusters. To this purpose, source and score data fusion techniques are used to achieve higher performance. A direct comparison shows that the experimental results using the proposed t-test feature selection and data fusion-based approach indicate superior performance when compared to classifiers that use all raw features and a data reduction method involving principal component analysis (PCA). In summary, the aim of this chapter is to introduce a novel and automatic statistical feature selection method based on the combination of t-test feature ranking and the Fisher Criterion of the VOI, which can be considered a lower-dimensional feature vector representation

of sMRI. The dimensionality of the feature vector can be adjusted by maximizing the Fischer Criterion in the training data-set. The proposed feature selection method not only selects the top discriminative features but also reduces the dimensionality of the input vectors to feature vectors. In addition, data fusion techniques are used to improve the AD classification performance among gray matter atrophy clusters. The performance of the proposed system is tested on 136 subjects (including 68 AD and 68 HC) from an ADNI dataset using 10-fold cross validation. The experimental results, when compared to those obtained with state-of-the-art techniques, show that the proposed system is highly competitive in terms of accuracy (96.32%), specificity (98.52%), and AUC (99.93%) for AD classification.

5.2 Material

5.2.1 Subjects

The diagnostic classification is conducted by selecting a total of 136 subjects from the ADNI database and grouping them as AD and HC. The AD group contained 68 subjects ranging in age from 61.4 to 89.2 (74.33±6.41) years. The Mini Mental State Examination (MMSE) and Clinical Dementia Ratio (CDR) scores ranged from 15 to 25 (mean 22.83±2.65) and 0.5 to2 (mean 0.75±0.41), respectively. The HC group contained 68 healthy controls ranging in age from 60.8 to 84.4 (74.14±4.95) years. The MMSE ranged from 28 to 30 (mean29.38±0.71) and the CDR is zero. A direct comparison revealed that the AD patients' mean MMSE and CDR are significantly distinct when compared to the HC subjects. No significant group differences are noted in age or sex ratio. Details of the demographics and clinical characteristics of the sample used in this chapter are presented in Table 5.1.

Table 5.1: Demographic and clinical details of the patients with AD and HC subjects

	AD (n=68)	**HC (n=68)**	**t-value**	**M.D**
Age	74.33±6.41	74.14±4.95	0.19	$0.18^{N.S}$
MMSE	22.83±2.56	29.38±0.71	14.76	-6.5^{*}
CDR [0/0.5/1/2]	0.75±0.41 [0/44/19/5]	0.0±0 [68/0/0/0]	-20.26	0.75^{*}

Note: All data present in mean ± standard deviation mode. AD, Alzheimer's Disease patients; CDR, Clinical Dementia Rating; HC, Healthy Control patients; MMSE, Mini-Mental State Examination; MD, Mean Difference; NS, Non-Significant; * , $p<0.0001$.

5.3 Proposed AD Classification System

This section proposed a new AD classification system using a novel approach based on a combination of t-test feature ranking and the Fisher Criterion for the optimal selection of feature vectors for high performance MRI classification of AD. The system involves five levels of processing. The pipeline of the proposed system is illustrated in Figure 5.1. First, the VBM plus DARTEL approach is employed to perform pre-processing on 3D MRI data. Second, a feature extraction method is used, based on VBM plus DARTEL analysis. Third, the extracted features are ranked based on the t-test values of the respective features, in the training set. In addition, an automatic approach based on the Fisher Criterion is adopted to determine the number of top ranking features. This approach adaptively determines the optimum number of top features and identifies a discriminative subset of high performance features based on training data in each fold. Hence, the feature vectors taken from VOIs of high dimensional s-MRI data are reduced into a low dimensional space, with improved discrimination capability. Fourth, the proposed technique is evaluated using state-of-the-art SVM classifiers. The performance analysis comprises an experimental setup based on 136 samples from the ADNI dataset. A 10-fold cross validation is employed throughout the performance analysis, which implies having 122 (90%)

samples in the training and 14 (10%) samples in the testing processes in each iteration. Finally, data fusion techniques among atrophy clusters are engaged to improve the classification performance.

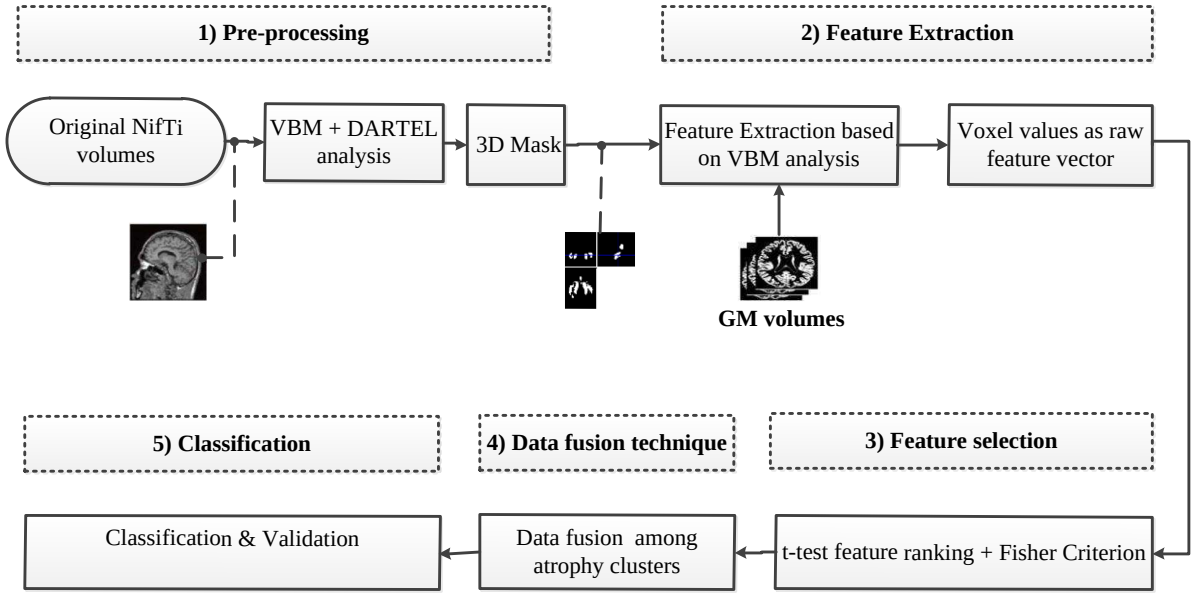


Figure 5.1: The pipeline of proposed system for classifying AD

5.3.1 MRI data preprocessing and statistical analysis

The MR images are pre-processed using the SPM8 and the VBM8 toolbox. VBM is an automated technique for assessment of the whole brain structure with voxel-by-voxel comparisons, developed to analyze tissue concentrations or volumes between subject groups for distinguishing degenerative diseases with dementia (J Ashburner & Friston, 2000; Nakatsuka et al., 2013). In more detail, VBM techniques investigate structural differences in areas with poorly defined structural landmarks (e.g., prefrontal areas) and provide explorative analysis of structural differences (John Ashburner, 2009; Cousijn et al., 2012; Takao, Hayashi, & Ohtomo, 2015). Recently, VBM has been applied to detect early atrophic changes in AD (Chételat et al., 2005; Hirata et al., 2005; Karas et al., 2003; Matsuda et al., 2012). It can provide statistical

results for comparisons of patients with AD and HCs (Matsuda et al., 2012). The inter-subject alignment of the MRI images is increased by applying the DARTEL approach, which has been reported to optimize the sensitivity of this type of analysis over standard VBM by using the Levenberg-Marquardt strategy (John Ashburner, 2007, 2009; Kasahara, Hashimoto, Abo, & Senoo, 2012; Klein et al., 2009; Modi et al., 2012). Moreover, the VBM8 toolbox benefits from the unified segmentation model with a maximum a posterior (MAP) technique (Rajapakse, Giedd, & Rapoport, 1997) and partial volume estimation (PVE) to account for partial volume effects (Tohka, Zijdenbos, & Evans, 2004), which results in a more subtle segmentation of subcortical areas. In addition, the VBM toolbox uses a spatially adaptive nonlocal means (SANLM) filter for denoising and removal of MRI in homogeneities (Manjón, Coupé, Martí-Bonmatí, Collins, & Robles, 2010). The signal-to-noise ratio is improved by employing a spatial constraint based on a classical Markov random field (MRF) model (Cuadra, Cammoun, Butz, Cuisenaire, & Thiran, 2005). Registration to a standard MNI-space (http://www.mni.mcgill.ca/) consists of a linear affine transformation and a nonlinear deformation using high-dimensional DARTEL normalization(John Ashburner, 2007). In the current work, sample homogeneity prior to calculating 2nd level analyses is ensured by inspecting the quality of gray matter images using the VBM8 toolbox. All MR images are corrected for bias field in homogeneities and then they are normalized and segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). The normalized and segmented images are modulated using a nonlinear deformation. In this work, only GM images are used. Finally, the 8 mm full-width-half-maximum (FWHM) Gaussian kernel is used for spatial smoothing of the GM images. After spatial pre-processing, the normalized, smoothed, modulated,

DARTEL-warped gray matter datasets are analyzed using a voxel-wise parametric mapping. The absolute threshold masking of around 0.1 is used to avoid possible edge effects around the border between gray matter and white matter or CSF. The regional gray matter volume changes are generated by voxel-based analysis over the whole brain. The framework of the general linear model is employed to detect gray matter volume changes in patients with AD using voxel-wise two sample t-test in SPM8. Age is engaged into the matrix design as a nuisance variable. The whole brain analysis is implemented using significance set at a p value of < 0.01, with correction for family-wise error (FWE) and a minimum cluster size of 1400 voxels for two-sample comparisons. Between-group differences in demographics and clinical parameters among or between groups of this work are evaluated using an independent two-sample t-test with the SPSS 16.0 package. (http://www.spss.com/). $p < 0.05$ is set as the level of significance.

5.3.2 Feature extraction

The feature extraction procedure based on VBM plus DARTEL analysis is applied to isolate the VOIs. The brain regions that show significantly decreased gray matter volumes, obtained using VBM plus DARTEL analysis, in AD patients relative to HC are segmented using 3D masks. For the segmented regions, the MarsBaR region of interest toolbox is employed (http://marsbar.sourceforge.net/) to generate cluster-specific binary masks. The center coordinates of each mask are defined by the local maximum revealed by VBM plus DARTEL analysis on the whole brain. These masks are applied to all the smoothed gray matter density volumes resulting from the VBM plus DARTEL analysis, to extract voxel values as raw feature vectors.

5.3.3 Feature selection

The dimensionality of raw feature spaces in the VBM extracted s-MRI voxel features is very high in comparison to the number of samples. The feature vectors span a very small region in the high dimensional vector space; consequently, a feature selection mechanism is desired in the post-processing. Feature selection can be considered in the form of a standard dimensionality reduction via a standard method, such as PCA. Alternatively, feature selection can be considered in the form of choosing the most discriminative subset of the available features in the raw feature vector. In this context, the proposed method can be employed, as it is the combination of t-test feature ranking and the Fisher Criterion, which not only reduces the dimensionality, but also increases the discriminability.

5.3.3.1 PCA dimensionality reduction

Principal component analysis is a statistical dimensionality reduction method that extracts a set of orthogonal principal components (PCs) from an original dataset (Haq et al., 2015; Lihua Tang et al., 2013). In this work, a 10-fold cross validation is used for measuring the performance of the classifiers. With136 samples, a 10-fold cross validation implied having 122 PCs through the PCA process. The number of PCs, h, used to generate the projection vectors of the training and testing data is chosen as h=122.

5.3.3.2 The general framework of feature ranking

The aim of feature ranking is to measure the relevance of features and class variables to aid in the selection of the most informative/discriminative features, thereby speeding up the learning process and promoting the performance of classifier models, especially when the dimensionality of the datasets is very large (N. Zhou & Wang, 2007). Let $D=\left[X_1, X_2, ..., X_N\right]^T$ be a dataset containing N samples, where

$X_i = (x_{i1}, x_{i2}, ..., x_{iM})$ is a vector of M values and each value x_{ij} of this vector shows a feature of that sample. The vector $f_j = (x_{1j}, x_{2j}, ..., x_{Nj})^T$ is a vector of values of a feature f_j. On the other hand, D represents a $N \times M$ matrix, where row i is the subject X_i and each column j is the feature f_j. A feature-ranking algorithm applied to dataset D generates an ordered list of the features $\Psi = [f_*^1, f_*^2, ..., f_*^j]$, where the superscript denotes the position in the ranked list of a feature f_* and this list is ordered by reduction importance. Based on feature ranking, we can select the top k ranked features $[f_*^1, f_*^2, ..., f_*^k]$, $k \leq M$ where k can be determined by the user or adjust experimentally(Prati, 2012). In this chapter, we use t-test feature-ranking approach, as follows (Kamkar, Gupta, Phung, & Venkatesh, 2014):

$$TS = \frac{\mu_{c1} - \mu_{c2}}{\sqrt{\frac{\sigma_{c1}^2}{n_{c1}} + \frac{\sigma_{c2}^2}{n_{c2}}}} \tag{5.1}$$

where TS is the t-test value and $\mu_{c1}, \sigma_{c1}^2, n_{c1}$ and $\mu_{c2}, \sigma_{c2}^2, n_{c2}$ are the mean, variance values, and number of samples of two classes c_1 and c_2, respectively. The top informative features are selected by ranking all features according to their TS values.

5.3.3.3 Optimal number of features based on Fisher Criterion

In addition to the feature-ranking algorithm based on the discriminative performance of the features, we propose the use of an automatic approach based on the Fisher Criterion, $J(w)$, given in Eq.(5.2), to determine the number of top discriminative features, thereby reducing the dimensionality of the prospective feature vectors (Diaf, Boufama, & Benlamri, 2013; Gao et al., 2012).

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{5.2}$$

Where S_B and S_W represent the determinants of between class and within class scatter matrices, respectively. For two classes c_1 and c_2, the between class scatter and within class scatter matrixes are defined as:

$$S_B = (\mu_{c1} - \mu_{c2})(\mu_{c1} - \mu_{c2})^T \tag{5.3}$$

$$S_W = \sum_{x_i \in c_1} (x_i - \mu_{c1})(x_i - \mu_{c1})^T + \sum_{x_i \in c_2} (x_i - \mu_{c2})(x_i - \mu_{c2})^T \tag{5.4}$$

Where $w = S_W^{-1}(\mu_{c1} - \mu_{c2})$ and μ_{ci} is the mean of data in each class. This approach helps in adaptively determining the k top discriminative features based on ranked t-test values using training data in each fold instead of using a fixed k. Once the features are ranked, the number of top ranked features iteratively increases from 1 to M (number of features) by calculating the respective Fisher Criterion. The number of top ranked features maximizing the Fischer Criterion is selected to be the optimal number of top ranked features k.The framework of the proposed feature selection method is illustrated in Figure 5.2.

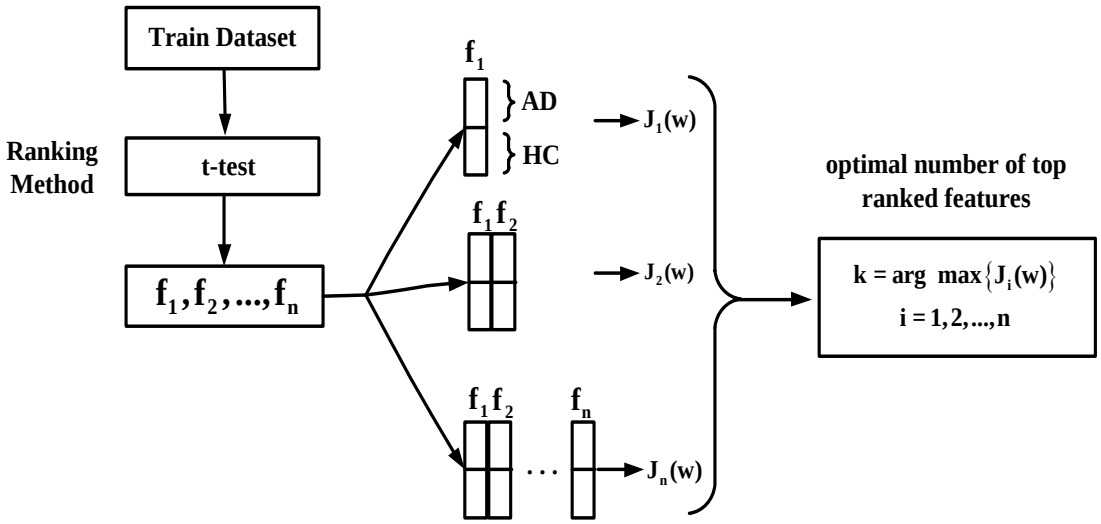


Figure 5.2**:** Schematic representation of proposed feature selection approach

5.3.4 Data fusion among atrophy clusters

This chapter introduces data fusion technique among atrophy clusters (VOIs) to improve the performance of the proposed AD classification method. The aim of the data fusion technique is to integrate the data from two or more distinct multiple sources (vectors, classifiers) to improve performance. In the current work, two different fusion techniques are used: source fusion and score fusion.

5.3.4.1 Source data fusion

In the scheme of source data fusion, the top features selected based on our approach, described in section 3.3, from different VOIs, are concatenated into a single feature vector. Assuming $fv_1, fv_2,..., fv_n$ are feature vectors generated using proposed feature selection method for each atrophy cluster. The feature vector fusion (FVF) is then:

$$fvf = \left[fv_1, fv_2,..., fv_n \right]_{1\times\sum_{i=1}^{n} m_i} \tag{5.5}$$

where m_i is the vector length for fv_i. This concatenated feature vector is then used for classification. The source data fusion relies on procedures for feature contraction.

5.3.4.2 Score data fusion

Score data fusion includes multiple classifiers and a combination method. The number of classifiers is determined based on the number of atrophy clusters obtained using the VBM plus DARTEL approach in the pre-processing. In this work, the majority voting method is employed as the score data fusion technique. Majority voting is one of the most versatile combination methods, because of its simplicity and performance on real data(Narasimhamurthy, 2005). The adopted score data fusion framework is illustrated in Figure 5.3.

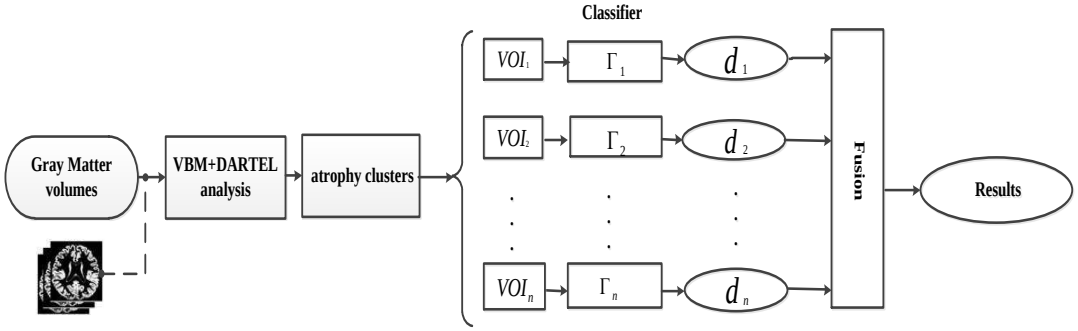


Figure 5.3**:** Majority voting based score data fusion

5.4 Experimental results and discussion

This section considers the experimental results obtained through the pre-processing phase using VBM plus DARTEL analysis on 3D T1weighted MR imaging, as an indicator disclosing the significance of decreased gray matter volumes in AD contributing to VOIs. The performance of the proposed feature selection method based on t-test ranking and the Fisher Criterion is also measured. Finally, the performance results obtained through data fusion are presented and analyzed. The performance of the classification using SVM classifiers with 10-fold cross validation is reported for the following cases: 1) performance of raw feature vectors directly extracted from VBM, 2) performance of the PCA data reduction method, 3) performance of proposed of t-test feature-ranking technique using the optimal

number of top features based on the Fisher Criterion, 4) performance of the proposed data fusion techniques among atrophy clusters of GM. The ACC (%), SEN (%), SPE (%) and AUC (%) performance metrics are used for the performance assessment.

5.4.1 Differences in gray matter volume between ADs and HCs

The gray matter volume atrophy differences between patients who suffer from AD and HC are summarized in Table 5.2. The group comparison by VBM plus DARTEL reveals a significant decline in GM volume in the right hippocampus (Talairach coordinates 26,-11,-9,x,y,z;z = Inf), left hippocampus (-25,-15,-8,x,y,z;z = Inf), right inferior parietal lobule (55,-44,25,x,y,z;z = 7.22), and right anterior cingulate (8,42,2,x,y,z;z = 6.54) (see Table 5.2 and Figure 5.4 for more details) in patients with AD when compared to the HCs. Figure 5.5 illustrates six three-dimensional views of group comparison representing relative gray matter atrophy in patients with AD compared to HCs. The voxel location of the significant atrophy regions are used as 3D VOI masks. These3D VOI masks are applied to the gray matter density volume results from the segmentation step in the VBM plus DARTEL analysis in order to extract voxel values into raw feature vectors for use in feature selection and classification. Based on these atrophy clusters, we define five different VOIs as follows:

VOI_1 includes the right hippocampus and amygdala regions. The center of this mask is at Talairach coordinates x=26, y=-11, z=-9. VOI_1 contains 16069 voxel values as a raw feature vector.

VOI_2 includes the left hippocampus-lateral globus pallidus regions. The center of this mask is at Talairach coordinates x=-25, y=-15, z=-8. VOI_2 contains 16974 voxel values as a raw feature vector.

VOI_3 includes the right inferior parietal lobule regions. The center of this mask is at Talairach coordinates x=55, y=-44, z=25. VOI_3 contains 1454 voxel values as a raw feature vector.

VOI_4 includes the right anterior cingulate regions. The center of this mask is at Talairach coordinates x=8, y=42, z=2. VOI_4 contains 2032 voxel values as a raw feature vector.

VOI_{all} includes all regions of gray matter loss (atrophy). VOI_{all} contains all four clusters above, with 36529 voxel values as a raw feature vector.

Note that the center of the mask in the Talairach coordinates corresponds to the center of the mass of the respective 3D VOI.

Table 5.2: Clusters of gray matter atrophy (68 AD vs. 68 HC)

Location of peak voxels	Hemisphere	Cluster size (no of voxels)	Talairach coordinates (x,y,z)			Z value (peak voxel)	T value (peak voxel)
Hippocampus-Amygdala	R	16069	26	-11	-9	Inf	10.94
Hippocampus-lateral globuspallidus	L	16974	-25	-15	-8	Inf	10.36
Inferior Parietal Lobule	R	1454	55	-44	25	7.22	8
Anterior Cingulate	R	2032	8	42	2	6.54	6.54

Note: Anatomical regions are derived from the Talairach Client program; L, heft hemisphere; R, right hemisphere; (FWE-corrected at $p < 0.01$).

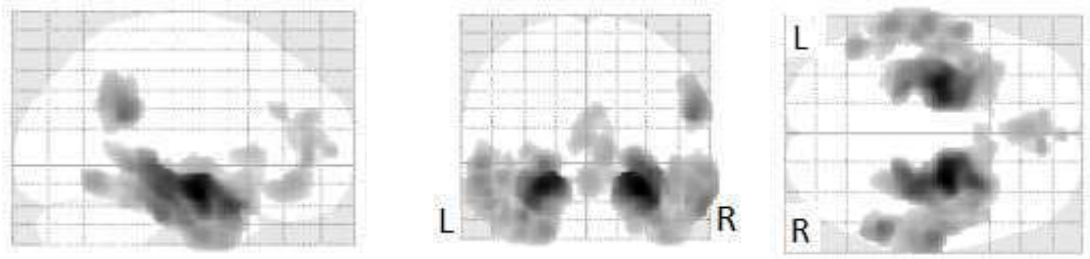


Figure 5.4:Brain regions where there are significant gray matter reduction (atrophy) in 68 patients with AD and 68 age matched HC subjects (FWE corrected at P < 0.01 and extend threshold K = 1400)

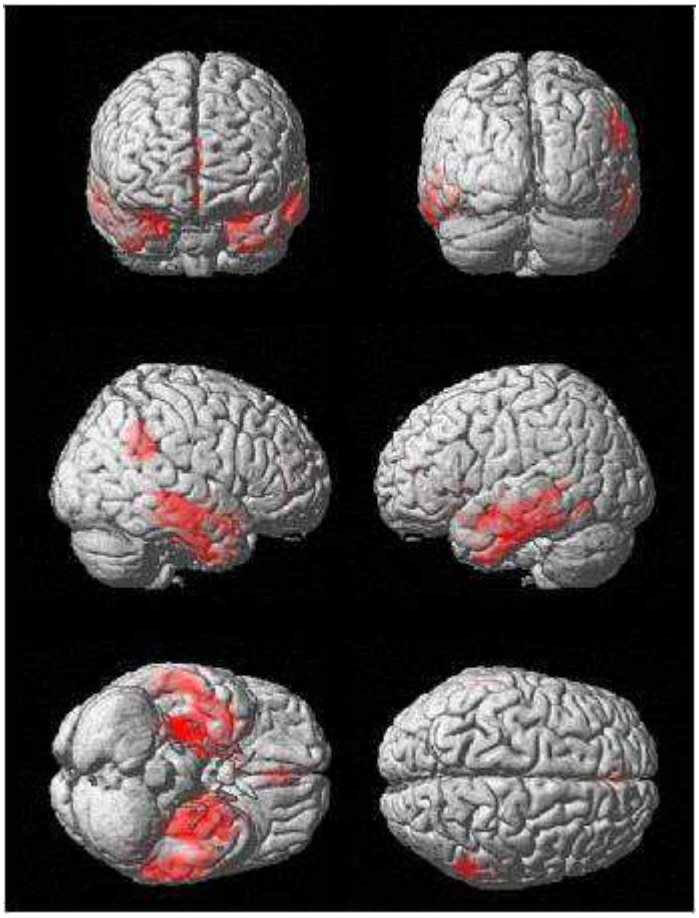


Figure 5.5: Three-dimensional reconstruction of the brain showing gray matter atrophy using VBM technique plus DARTEL. The regions of gray matter loss are shown from anterior, posterior, right lateral, left lateral, inferior and superior view, respectively. The red region represents the region of gray matter loss

5.4.2 Performance of the raw feature vectors

The complete MRI dataset consists of 68 AD and 68 HC samples. The ACC, SEN, SPE, and AUC that are obtained by 10-fold cross validation using the SVM classifier (Linear and RBF kernels) on raw feature vectors from five different VOIs are shown in Table 5.3. The results indicate that the average performance, in terms of ACC, SEN, SPE, and AUC obtained from five atrophy clusters using RBF SVM, is marginally better than Linear SVM. The RBF kernel is generally more flexible than the linear kernel so it generally can model more functions with its function space.

Table 5.3: Raw feature vectors performance of atrophy clusters using 10 fold cross validation

	Linear SVM				RBF SVM			
	ACC(%)	**SEN(%)**	**SPE(%)**	**AUC(%)**	**ACC(%)**	**SEN(%)**	**SPE(%)**	**AUC(%)**
VOI_1	80.14	79.41	80.88	85.37	82.35	80.88	83.82	88.71
VOI_2	77.20	77.94	76.47	84.93	79.41	76.47	82.35	87.69
VOI_3	71.32	70.58	72.05	75.65	75.00	72.05	77.94	80.75
VOI_4	69.85	69.11	70.58	77.82	70.58	73.52	67.64	77.99
VOI_{all}	77.20	79.41	75.00	84.49	83.82	83.82	83.82	86.00
Average	75.14	75.29	74.99	81.65	78.23	77.34	79.11	84.22

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; SVM, Support Vector Machine; RBF, Radial Basis Function.

5.4.3 Performance of the PCA method

The PCA based data reduction method is utilized to extract raw feature vectors. For each dataset, the features extracted are reduced to lower dimensional features using PCA, with 122 PCs. Table 5.4 presents the classifier performance obtained using 10-fold cross validation for SVM classifiers in terms of ACC, SEN, SPE, and AUC. The average accuracy of feature vectors with 122 PCs for linear and RBF SVM classifiers is74.20% and 78.08%, respectively, while the average accuracy using the raw feature vectors without dimensionality reduction is 75.14% and 78.23%, respectively. As observed, PCA introduces dimensionality reduction and generates comparable performance with the raw data.

Table 5.4: PCA performance of atrophy clusters using 10 fold cross validation with 122 PCs

	Linear SVM				RBF SVM			
	ACC(%)	**SEN(%)**	**SPE(%)**	**AUC(%)**	**ACC(%)**	**SEN(%)**	**SPE(%)**	**AUC(%)**
VOI_1	79.41	82.35	76.47	86.80	81.61	86.76	76.47	88.27
VOI_2	74.26	76.47	72.05	83.06	82.35	82.35	82.35	87.59
VOI_3	70.58	73.52	67.64	72.48	69.85	69.11	70.58	78.33
VOI_4	69.58	69.11	70.58	80.54	71.32	66.17	76.47	79.35
VOI_{all}	77.20	79.41	75.00	87.49	85.29	86.47	83.82	88.74
Average	74.20	76.17	72.34	82.07	78.08	78.17	77.93	84.45

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; SVM, Support Vector Machine; RBF, Radial Basis Function.

5.4.4 Performance of the proposed feature selection using t-test ranking and the Fisher Criterion

As proposed in section 5.3.3.3, the feature selection technique uses the t-test for ranking the features. The Fisher Criterion is used to determine the optimal number of top features. The Fisher scores for the samples in the training set from fold 1 of VOI_{all} are plotted in Figure 5.6 for the top 250 ranked features. As Figure 5.6 shows, the maximum Fisher score is located at 111, which means that 111 top-ranked features are to be used in the classification process. Typical Fisher scores are observed between 30 and 150 for all folds of 5 different VOIs. Figure 5.7 shows all of the t-test values for the same data. The contribution of features on the accuracy is studied separately and plotted in Figure 5.8 with linear SVM. As expected, the contribution of the features in relevance to their t-test values is highly correlated. A higher t-test rank implies higher performance of the respective feature. A logarithmic scale is used to cover the entire feature space. Additionally Figure 5.9 is included to show the improvement in the accuracy obtained by using progressive inclusion of the

ranked features in the feature vector with linear SVM. The performance increases with the increased number of ranked features used in the classification. However, after a certain maximum, which corresponds to 111 top ranked features in this fold, the performance does not increase further. The SVM-based classifiers are used to observe the classification performance of the selected feature vectors from five different VOIs. The results of classifiers are presented in Table 5.5. Examination of Tables 5.3 and 5.5 confirms that the proposed feature section method significantly improves the prediction capability of AD subjects when compared to prediction using raw features. The average accuracy for raw data for linear and RBF SVM classifiers is 75.14% and 78.23%, respectively, while the average accuracy for the proposed feature selection method is 86.76% and 86.76%, respectively. The improvement is around 10% for all performance indicators: ACC, SEN, SPE, and AUC.

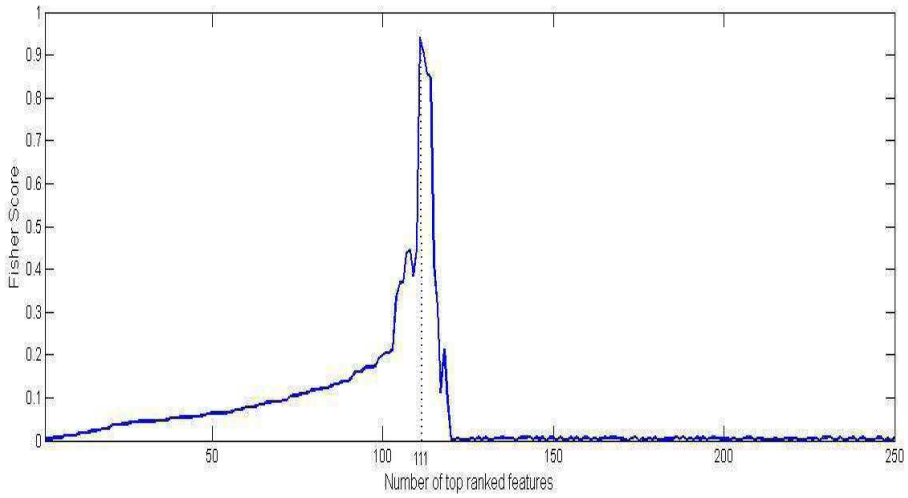


Figure 5.6: Fischer scores for the respective ranked features in fold 1 training of VOI_{all}

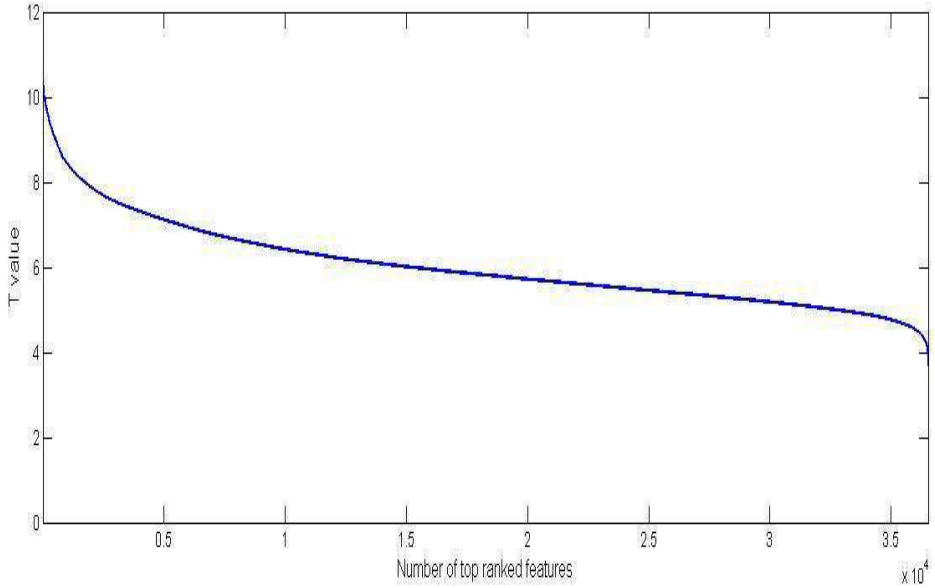


Figure 5.7: t-test (TS) values for the respective ranked features in fold 1 training of VOI_{all}

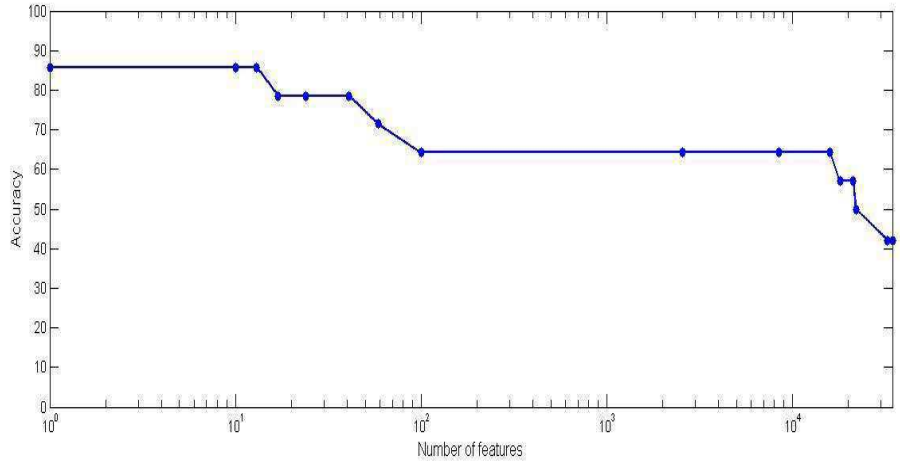


Figure 5.8: Classification accuracies of linear SVM with respect to different numbers of features selected in fold 1 training of VOI_{all}

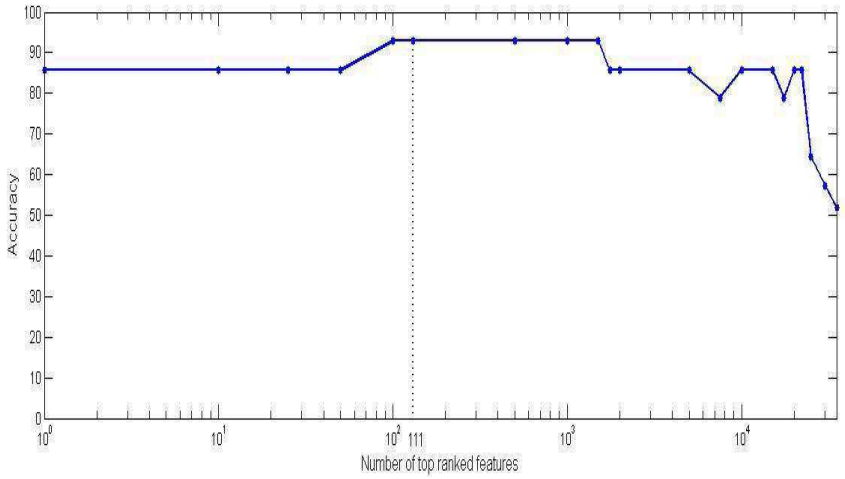


Figure 5.9: Classification accuracies of linear SVM with respect to different numbers of top ranked features selected in fold 1 training of VOI_{all}

Table 5.5: Performance results of the proposed feature selection method

	Linear SVM				RBF SVM			
	ACC(%)	**SEN(%)**	**SPE(%)**	**AUC(%)**	**ACC(%)**	**SEN(%)**	**SPE(%)**	**AUC(%)**
VOI_1	91.17	92.64	89.70	96.9	90.44	89.70	91.17	95.07
VOI_2	92.64	91.17	94.11	97.93	94.11	92.64	95.58	98.74
VOI_3	76.47	73.52	79.41	84.93	76.47	75.00	77.94	84.66
VOI_4	79.41	75.00	83.82	86.67	80.14	73.52	86.76	89.29
VOI_{all}	94.11	95.58	92.64	98.33	92.64	94.11	91.17	98.13
Average	86.76	85.58	87.93	92.95	86.76	84.99	88.52	93.17

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; SVM, Support Vector Machine; RBF, Radial Basis Function.

5.4.5 Performance of data fusion among atrophy clusters

The performance improvement aided by data fusion of five clusters is shown in Table 5.6. The performance of both types of data fusion techniques is around 10% higher than the average performance obtained with individual clusters. The performance of the majority voting (score fusion) approach is always higher than or equal to the performance of the source concatenation (source fusion) approach. Table

5.6 shows that data fusion among atrophy clusters of GM volumes integrates information by improving the classification performance in all terms.

Table 5.6: Performance of proposed data fusion technique among atrophy clusters of GM

	Linear SVM				RBF SVM			
	ACC(%)	**SEN(%)**	**SPE(%)**	**AUC(%)**	**ACC(%)**	**SEN(%)**	**SPE(%)**	**AUC(%)**
Source Concatenation	95.58	94.11	97.05	97.52	95.58	94.11	97.05	97.31
Majority Voting	**96.32**	**94.11**	**98.52**	**99.93**	95.59	94.11	97.05	99.82

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; SVM, Support Vector Machine; RBF, Radial Basis Function.

5.5 Performance comparison to the other methods

Several recent studies have reported classification results to distinguish AD and HC based on MRI. Zhang et al.(D. Zhang et al., 2011) used multimodal classification of AD based on the combination of MRI, CSF, and PET. They reported an ACC of 86.2% in the classification of AD/HC by MRI image modality. They also achieved a high ACC performance of 93.2% by combining the MRI, CSF, and PET results. Westman et al.(Westman et al., 2012) reported an ACC of 87% from MRI data and increased it to 91.8% by combining MRI data with CSF measures. Zhou et al. (Q. Zhou et al., 2014)employed FreeSurfer software to calculate 55 volumetric variables from MRI. They reported an ACC of 78% for MRI data and 92.4% for combining MRI data with MMSE. In the present work, only the MRI modality with 136 samples from the ADNI dataset is used, with highly comparable results to those reported in other MRI-only studies. The performance of the proposed feature selection and data fusion techniques outperforms the alternative techniques are given in Table 5.7. The detail parameters of classification performance with different methods on MRI data are also provided in Table 5.7. The results reported in Table 5.7 show that the

performance of the proposed system is highly competitive for the performance terms including ACC, SPE, and AUC when compared to the other systems reported in the literature. The only exception is SPE, where the performance of the proposed system is lower than for results reported by Kloppel et al., 2008 (Klöppel et al., 2008) for groups I and II. Our results are highly competitive with the rest of the systems. The performance improvement over the previous work, shown in Table 5.7, can be attributed to the automatic statistical feature-selection method based on the combination of t-test feature ranking and the Fisher Criterion of the VOI. Due to t-test ranking, the proposed feature selection method is capable of sorting discriminative features in descending order. The optimal dimension is of the feature vector is adjusted by maximizing the Fischer Criterion in the training dataset. Finally, data fusion techniques among gray matter atrophy clusters provide further improvement on the AD classification performance.

Table 5.7: Supervised classification results of Alzheimer's disease and healthy control subjects on MRI data

Author	Imaging Modality	Source of data	AD/HC	Validation method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
Zhang et al.,2011(D. Zhang et al., 2011)	MRI	ADNI	51/52	10 Fold	86.2	86.0	86.3	-
Zhang et al.,2011(D. Zhang et al., 2011)	MRI + CSF + PET	ADNI	51/52	10 Fold	93.2	93.0	93.3	-
Westman et al., 2012 (Westman et al., 2012)	MRI	ADNI	96/111	10 Fold	87	83.3	90.1	93.0
Westman et al., 2012 (Westman et al., 2012)	MRI + CSF	ADNI	96/111	10 Fold	91.8	88.5	94.6	95.8
Zhou et al.,2014 (Q. Zhou et al., 2014)	MRI	Private	127/59	2 Fold	78.2	68.5	81.7	-
Zhou et al.,2014(Q. Zhou et al., 2014)	MRI + MMSE	Private	127/59	2 Fold	92.4	84.0	96.1	-
Kloppel et al., 2008 (Klöppel et al., 2008)	MRI (Group I)	Private	20/20	Leave-one-out	95.0	95.0	95.0	-
Kloppel et al., 2008 (Klöppel et al., 2008)	MRI (Group II)	Private	14/14	Leave-one-out	92.9	100	85.7	-
Kloppel et al., 2008 (Klöppel et al., 2008)	MRI (Group III)	Private	33/57	Leave-one-out	81.1	60.6	93.0	-
Hinrichs et al., 2011(Hinrichs et al., 2011)	MRI + PET	ADNI	48/66	10 Fold	87.6	78.9	93.8	-
Hinrichs et al., 2011(Hinrichs et al., 2011)	MRI + PET + CSF + APOE +Cognitive Scores	ADNI	48/66	10 Fold	92.4	86.7	96.6	-
Proposed method	**MRI**	**ADNI**	**68/68**	**10 Fold**	**96.32**	**94.11**	**98.52**	**99.93**

5.6 Conclusion

This chapter proposes a feature selection method using t-test-based feature ranking, which is used for the classification of AD. The optimal size of the selected features is determined using the Fisher Criterion, which maximizes the class separation between AD and HC. The feature selection is applied to all voxels that pass through masks modelled by overall atrophy clusters, determined by using VBM analysis. Linear and RBF kernel-based SVM classifiers are used for the classification of the extracted feature vectors after the proposed feature selection method. A performance improvement is also proposed by applying data fusion among the individual atrophy clusters, as well as the overall atrophy clusters. Standard data fusion techniques, such as source and score fusion, are used to obtain improved performance in the classification of AD. The performance of the proposed system is measured on 136 subjects (68 AD and 68 HC) from the ADNI dataset using 10-fold cross validation. The experimental results show that the performance of the proposed approach for ACC, SPE, and AUC is highly competitive with the state-of-the-art techniques using MRI data reported in the literature.

Chapter 6

STRUCTURAL MRI-BASED DETECTION OF ALZHEIMER'S DISEASE USING FEATURE RANKING AND CLASSIFICATION ERROR

6.1 Introduction

This chapter describes the application of an automatic CAD system, which uses statistical feature-ranking methods as part of a novel feature-selection process, followed by estimation of the classification error in AD and healthy control (HC) groups to determine the optimum number of highest-ranking features to be selected. In the training set, resubstitution and cross-validation error estimators are used as classification errors to measure the quality of a classifier. We used these classification error metrics as stopping criteria among the ranked features to estimate the optimal number of features with the most discriminative information in the classification process. We evaluated seven feature-ranking methods, namely, statistical dependency (SD),mutual information (MI), information gain(IG), Pearson's correlation coefficient (PCC), the *t*-test score (TS), Fisher's criterion (FC), and the Gini index (GI) in the proposed CAD system. In the proposed approach, high-dimensional feature space is reduced into lower dimensional space by employing the minimized classification error as the dimensionality selection criterion in an iterative process of incrementing the number of ranked features. The proposed feature-selection method is applied to gray matter (GM) atrophy clusters of voxels,

which corresponded to the volume of interests (VOIs) of the sMRI data obtained through the voxel-based morphometry (VBM) analysis during preprocessing. VBM is an advanced method used to assess the whole-brain structure using voxel-by-voxel comparisons (J Ashburner & Friston, 2000; Guo et al., 2010; Matsuda et al., 2012; Moradi et al., 2015; Nakatsuka et al., 2013). It is one of the best methods for feature extraction from sMRI in AD(Bron et al., 2015). In the proposed system, we used only sMRI data. The proposed CAD system is applied in four stages in a systematic manner. In the first stage, the VBM technique is employed, in addition to diffeomorphic anatomical registration using the exponentiated Lie algebra (DARTEL) (Matsuda et al., 2012). This approach is used to analyze group-wise comparisons between cross-sectional structural MRI scans to detect the MRI voxels that are best discriminated between the AD group versus HCs (J Ashburner & Friston, 2000; Matsuda et al., 2012; Moradi et al., 2015; Nakatsuka et al., 2013). Based on the VBM and DARTEL approach on a global brain scale, and regional structural GM alterations, regions with significant atrophy of GM are investigated and specified in the patients who suffer from AD. In the second stage, specified VOIs are used as 3D masks for extracting voxel intensity values from the VOIs to generate feature vectors. These feature vectors are subjected to further data-selection processes before they are used by the classifier. In the third stage, the extracted features are ranked based on the statistical scores (i.e., SD, MI, IG, PCC, TS, FC, and GI) of the AD and HC groups in the training set. The ranking scores can be considered an indicator of the level of separation/discrimination between the AD and HC groups in the training set. Feature ranking has been used successfully in a number of pattern-recognition studies (Chang & Lin, 2008; Duch, Wieczorek, Biesiada, & Blachnik, 2004; Geng, Liu, Qin, & Li, 2007; Prati, 2012; Ruiz,

Riquelme, & Aguilar-Ruiz, 2003; Slavkov, Zenko, & Dzeroski, 2010). In addition, an automatic approach based on classification error estimation is used to determine the number of top features using the AD and HC groups in the training set. This approach adaptively determines the optimum number of top features and identifies a discriminative subset of high-performance features based on the training data in each fold instead of using a fixed number of features. In the fourth stage, the performance of the proposed feature-selection technique is evaluated using a support vector machine (SVM) classifier. In this work, the SVM classifier with a linear kernel is trained to discriminate between the classes. In addition, instead of using a single feature ranking method, the results of multiple individual feature ranking methods are combined through the proposed data fusion technique for improved classification performance.

In summary, the aim of this chapter is to design an automatic CAD system based on statistical feature ranking and classification errors as part of a novel feature-selection method. The proposed system utilizes feature ranking based on statistical scores, followed by the determination of resubstitution and cross-validation error estimators to identify the number of ranked features that minimizes the error in the training set. This process helps to identify a selected discriminative subset of high-performance features into a lower-dimensional feature vector space representing sMRI images. In addition, a data fusion technique is proposed to improve the AD classification performance among different feature ranking methods. The performance of the proposed system is assessed using a data set from the Alzheimer's Disease Neuroimaging Initiative (ADNI) containing 260 subjects (130 AD patients and 130 HCs) using 10-fold cross-validation. The experimental results showed that the

accuracy (ACC) (92.48%), sensitivity (SEN) (91.07%), specificity (SPE) (93.89%), and area under the curve (AUC) (96.30%) of the proposed system are well comparatively to results obtained with state-of-the-art techniques in terms of AD classification.

6.2 Materials

6.2.1 MRI acquisition

The MR images and data used in this chapter are obtained from the ADNI database (www.loni.ucla.edu/ADNI). All the participants initially underwent a number of neuropsychological examinations, resulting in several clinical characteristic indicators, including the Mini Mental State Examination (MMSE) score and Clinical Dementia Ratio (CDR) score. The HC group contained 130 participants, with ages ranging from 56 to 88 years (mean 74.49±6.13 years), MMSE scores ranging from 27 to 30 (mean of 29.26±0.80), and a CDR score of zero. The AD group contained 130 patients, and their ages ranged from 57 to 91 years (mean of 75.88±7.54 years). Their MMSE and CDR scores ranged from 10 to 28 (mean of 22.33±3.27), and 0.5 to 2 (mean of 0.80±0.37).

6.3 Proposed CAD classification system

In this section, an automatic CAD system, which is based on feature ranking, followed by optimal selection of a number of top features using a classification error for high-performance AD classification, is introduced. An outline of the proposed ranking-based CAD system is illustrated in Figure 6.1. First, the VBM and DARTEL approach are employed to preprocess3D T1-weightedMRI data. Second, voxel-based feature extraction is performed. Third, the extracted features are ranked based on the score values of the respective features in the training set. The optimal number of top ranked features is automatically obtained by minimizing the classification error

among the possible number of features. These approaches resulted in high-dimensional sMRI data of VOI in a low-dimensional space with a discriminative subset of high-performance features based on the training data in each fold. Fourth, to evaluate the performance of the proposed feature-selection method, a linear SVM classifier is employed. In addition, a data fusion technique among different feature ranking methods is engaged to improve the classification performance.

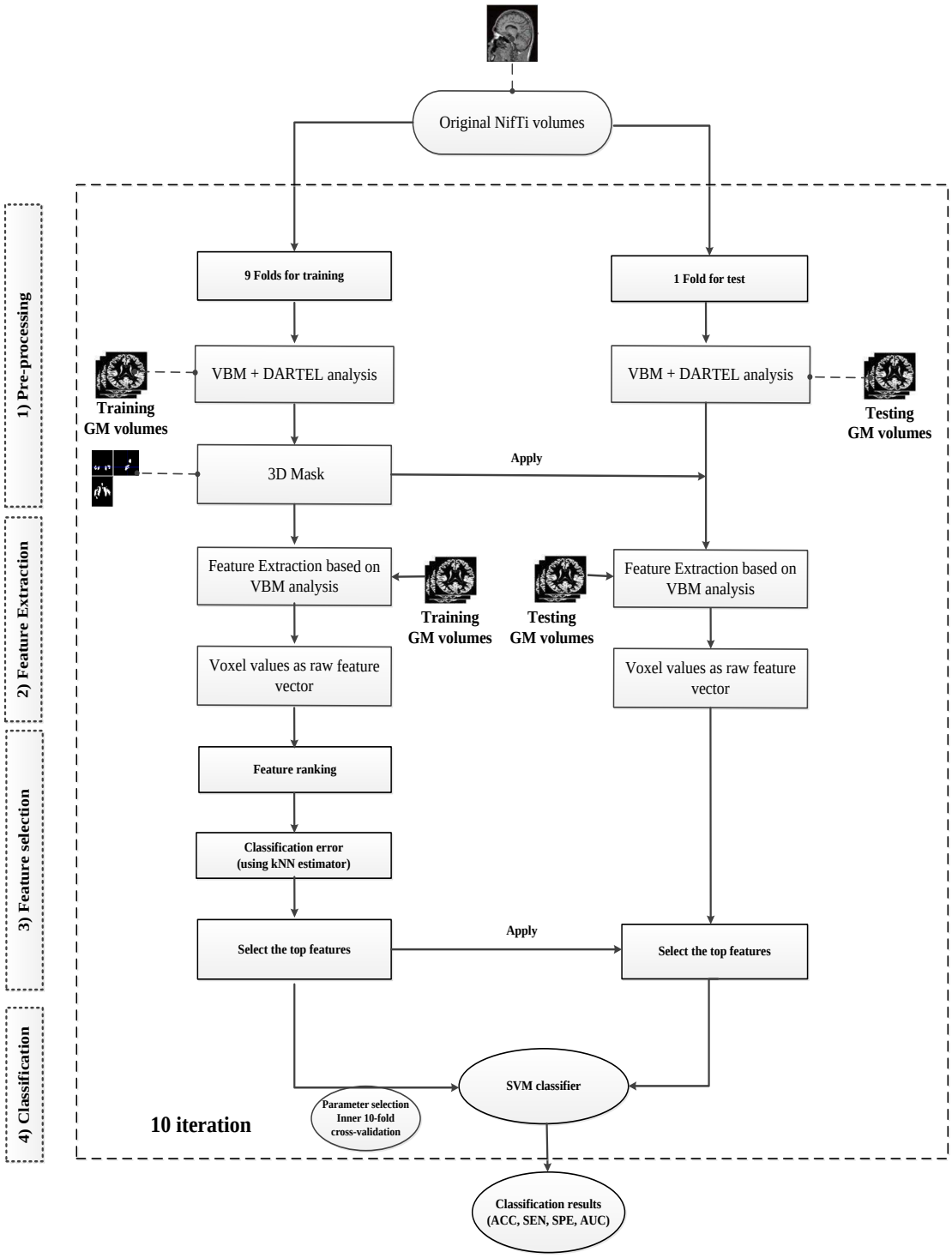


Figure 6.1: The pipeline of proposed ranking-based CAD system for classifying AD

6.3.1 MRI data preprocessing

The 3D T1-weighted brain images are pre-processed using the SPM8 package and VBM 8 toolbox. Recently, several studies have been used VBM method for detection

atrophic changes in AD (Bron et al., 2015; Hirata et al., 2005; Matsuda et al., 2012; Moradi et al., 2015; Son, Han, Min, & Kee, 2013; Xu, Wu, Chen, & Yao, 2015). In this study, DARTEL approach is employed with VBM to increase enhancement of inter-subject registration provide precise, accurate localization of structural damage of the MRI images. DARTEL template is generated from 550 healthy control participants (defined by default setting of VBM8 toolbox) (Cousijn et al., 2012). In the VBM8 toolbox, all the sMRI data are bias-corrected and segmented into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) components. The normalized segmented images are modulated by applying a nonlinear deformation. This allows the comparison of absolute amounts of tissue corrected for individual differences in brain size(Cousijn et al., 2012). The deformation is applied to segmented images to create an image which is in voxel-for-voxel registration with the template(Greve, 2011). In the present thesis, we used only GM component. Finally, the all GM components are spatially smoothed with an 8 mm full-width-half-maximum Gaussian smoothing kernel. After spatial preprocessing, the smoothed, modulated, DARTEL-warped and normalized GM datasets are subjected to a statistical analysis using a p-value of <0.01 with correction for family-wise error (FWE). The extent threshold is adjusted at 1,400 voxels for two-sample comparisons. Regional changes in GM volumes are detected by a voxel-based analysis of the entire brain.

6.3.2 Feature extraction

The brain regions containing significantly decreased GM volumes obtained using the VBM plus DARTEL analysis in the AD patients relative to the HCs. Based on the VBM and DARTEL results, a 3D mask is modeled to identify VOIs for further processing. This mask is applied to the GM density volumes resulting from the VBM

and DARTEL analyses to extract voxels as feature vectors. The subjects are randomly divided into 10 folds, with the same number of AD and HC subjects in each fold. In each iteration, one of the folds is used for testing, and nine of the folds are used for training. A VBM analysis of each training data set is performed to reveal regions of decreased GM volume in the patients through a 3D mask for the MRI samples in the respective training fold. In total, 10 different masks with different lengths (i.e., from 59395 to 69170 voxels) are defined. The respective 3D masks are used in the respective iteration to extract features from the training and testing data sets.

6.3.3 Proposed feature selection

The dimensionality of raw feature spaces, which is very high, changed in line with the dimensionality of the 3D masks (i.e., from 59395 to 69170 voxels). It is expected that the feature vectors span a smaller region in the high-dimensional vector space. The aim of feature selection is to select the best features for improving the efficiency of learning, component al cost and classification performance. Feature selection using feature ranking is a reasonable approach to reduce the dimensionality and improve the performance ,as the most discriminative subset of features are employed as the top features representing the samples. Recently, several studies have used different feature ranking methods as part of feature selection in pattern recognition field (Chang & Lin, 2008; Duch et al., 2004; Geng et al., 2007; Pohjalainen, Räsänen, & Kadioglu, 2015; Ruiz et al., 2003; Slavkov et al., 2010; W. Yan, 2007).

6.3.3.1 Feature ranking

Feature ranking aids to achieve knowledge of data and identify relevant features and sort the features with respect to their relevance. on other hand, feature ranking makes it easier to determine the relevance of features and class variables and to select the

most informative/discriminative features, thereby improving the performance of classifier models and speeding up the learning process, especially when the dimensionality of a data setis very large (N. Zhou & Wang, 2007). Let $\Phi = [f^1, f^2, ..., f^M]$ be a features set containing M features, where the vector $f^j = (x_1^j, x_2^j, ..., x_N^j)^T$ is a vector of the values of a feature, f^j, N is the number of samples, and each value x_i^j of this vector shows a feature of that sample. A feature-ranking algorithm applied to data set Φ generates an ordered list of the features $\Psi = [f_*^1, f_*^2, ..., f_*^M]$. The superscript denotes the position in the ranked list of a feature, f_*,and the list is ordered by the reduction importance. Based on the feature ranking, we can select the top q-ranked features $[f_*^1, f_*^2, ..., f_*^q]$ $q \leq M$, where q can be determined by the user or adjusted experimentally (Prati, 2012). In the present work, q is automatically estimated by minimizing the classification error of the training set in each fold. In the present thesis, we used the following seven feature-ranking approaches. In each approach, the score of each feature is computed independently and sorted based on the respective score.

1. SD: SD measures the level of dependency between the values of a feature and the associated class labels. The SD between feature value X and class label C can be obtained as follows (Pohjalainen et al., 2015) :

$$SD = \sum_i \sum_j P(\mathrm{x_i}, \mathrm{c_j}) \frac{P(\mathrm{x_i}, \mathrm{c_j})}{P(\mathrm{x_i})P(c_j)} \tag{6.1}$$

where $P(\mathrm{x_i}, \mathrm{c_j})$ is the frequency count of data X with value x_i in the class c_j, $P(\mathrm{x_i})$ is the frequency count of data X with value x_i, and $P(c_j)$ is the frequency count of class C with value c_j. SD is nonnegative in the range of [0, 1], with SD=0 indicating no correlation and SD=1 denoting that *C* can be inferred once *X* is known. A larger SD means higher dependency between the feature value and class labels.

2. MI: MI measures the relevance of the feature value *X* and class label *C* by (Cabral et al., 2015; Pohjalainen et al., 2015; W. Wang et al., 2014):

$$MI = \sum_i \sum_j P(\mathrm{x_i}, \mathrm{c_j}) \log_2 \frac{P(\mathrm{x_i}, \mathrm{c_j})}{P(\mathrm{x_i}) P(c_j)} \tag{6.2}$$

MI is similar to SD. $P(\mathrm{x_i}, \mathrm{c_j})$ is the frequency count of data X with value x_i in the class c_j, $P(\mathrm{x_i})$ is the frequency count of data X with value x_i, and $P(c_j)$ is the frequency count of class C with value c_j. MI is nonnegative in the range of [0, 1], with MI=0 indicating no correlation, and MI=1 meaning that *C* can be inferred once *X* is known.

3. IG: IG is a measure of the dependence between the features and class label. The IG of feature value *X* and class label *C* is calculated as follows(Zhao et al., 2010):

$$IG = H(X) - H(X \mid C) \tag{6.3}$$

where $H(X)$ and $H(X \mid C)$ are the entropy of *X* and the entropy of *X*, respectively, after observing *C*, as follows:

$$H(X) = -\sum_i P(x_i) \log_2 (P(x_i)) \tag{6.4}$$

$$H(X|C) = -\sum_{j} P(c_j) \sum_{i} P(x_i|c_j) \log_2(P(x_i|c_j)) \quad (6.5)$$

The maximum value of IG is 1. Features with higher IG are more relevant.

4. PCC: PCC is a measure of the relevance between the features and class label. PCC of the feature value *X* and class label *C* is calculated as follows (W. Wang et al., 2014):

$$PCC = \frac{\text{cov}(X, C)}{\sqrt{\text{var}(X)\,\text{var}(C)}} \quad (6.6)$$

which in binary classification becomes:

$$PCC = \frac{\sum_{i=1}^{N} (x_i - \mu_x)(c_i - \mu_c)}{\sqrt{\sum_{i=1}^{N} (x_i - \mu_x)^2 \sum_{i=1}^{N} (c_i - \mu_c)^2}} \quad (6.7)$$

Where *PCC* is Pearson's correlation value, and μ_x and μ_c are the mean of all samples of *X* and *C*, respectively. *PCC* has a value in the range of [-1, 1]. PCC=0 indicates independency of *X* and *C*, PCC=1 denotes the highest positive correlation of them, and PCC=-1denotes the highest negative correlation. To select the top informative features, all the features are ranked according to their absolute PCC values.

5. TS: The TS measures the statistical significance of the value differences between the two classes. The *t*-test is performed by(Kamkar et al., 2014):

$$TS = \frac{\mu_{c1} - \mu_{c2}}{\sqrt{\frac{\sigma_{c1}^2}{n_{c1}} + \frac{\sigma_{c2}^2}{n_{c2}}}} \quad (6.8)$$

where TS is the *t*-test value and μ_{c1}, σ_{c1}^2, n_{c1} and μ_{c2}, σ_{c2}^2, n_{c2} are the mean, variance values, and number of samples of two classes, c_1 and c_2. To select the

top informative features, all the features are ranked according to their absolute TS values.

6. FC: FC measures between-class and within-class scatter matrices between two classes, as shown below:

$$FC = \frac{w^T S_B w}{w^T S_W w} \tag{6.9}$$

where S_B and S_W represent the determinant of the between-class and within-class scatter matrices, respectively(Gao et al., 2012). For two classes c_1 and c_2, the between-class scatter and within-class scatter matrixes are defined as follows:

$$S_B = (\mu_{c1} - \mu_{c2})(\mu_{c1} - \mu_{c2})^T \tag{6.10}$$

$$S_W = \sum_{x_i \in c_1} (x_i - \mu_{c1})(x_i - \mu_{c1})^T + \sum_{x_i \in c_2} (x_i - \mu_{c2})(x_i - \mu_{c2})^T \tag{6.11}$$

where $w = S_W^{-1}(\mu_{c1} - \mu_{c2})$ and μ_{ci} are the mean of the data in each class.

To select the top informative features, all the features are ranked according to their FC values.

7. GI: The GI is a measure used to quantify the ability of a feature to distinguish between classes. The GI for a feature, *f*, is as follows(Zhao et al., 2010):

$$GI(f) = 1 - \sum_{i=1}^{C} \left[p(i \mid f \right]^2 \tag{6.12}$$

In the binary classification, the maximum value of the GI is 0.5, and features with a smaller GI are more relevant.

6.3.3.2 Classification error

Consider a labeled feature vector, $D = \{X, T\}$, where $X \in \Re^p$ (p is the dimension of the input vector) and T is the class label, which in binary classification with two classes $T \in \{-1, 1\}$ .The pair $\{X, T\}$ has a joint probability distribution, F ,which is

unknown in practice. Let a classifier be trained with a set of n -independent observations, $S_n = \{(\mathrm{x}_1, t_1), ..., (\mathrm{x}_\mathrm{n}, \mathrm{t}_\mathrm{n})\}$, which are drawn from F . Let $\varphi : \{\Re^p \times \{-1,1\}\}^n \times \Re^p \to \{-1,1\}$ be a mapping input space to target as a classification rule, which maps S_n onto a classifier, $\varphi_n : \Re^p \to \{-1,1\}$ (Sima, Braga-Neto, & Dougherty, 2011). The classification error e_n is the probability of an erroneous classification, which is calculated as follows (Sima, Braga-Neto, & Dougherty, 2005; Sima et al., 2011):

$$e_n = P(\varphi_n(X) \neq T \,|\, \mathrm{S}_\mathrm{n}) \tag{6.13}$$

In practice, the classification error is unknown, and the error must be estimated ($\hat{e}$).In the present work, two different classification error estimators are used: a resubstitution error ($\hat{e} = e_{resub}$) and across-validation error ($\hat{e} = e_{cross}$).

6.3.3.2.1 Resubstitution error

Consider a classifier, φ , which is trained with a set, $S_n = \{(\mathrm{x}_1, t_1), ..., (\mathrm{x}_\mathrm{n}, \mathrm{t}_\mathrm{n})\}$, where n is the number of samples. In the resubstitution error, e_{resub} , we design a classifier, S_n and test it on S_n to estimate the respective error, as follows:

$$e_{resub} = \frac{1}{n} \left\| (T_i - \varphi_n(x_i)) \right\|_0 \tag{6.14}$$

Where $i = 1, ..., n$, and $\|v\|_0$ is the zero-norm counting the number of nonzero entries in v. The resubstitution estimator is nonrandomized, and it is very fast to compute in comparison to other error estimators, such as the cross-validation error estimator (Braga-Neto, 2009). This estimator is always optimistically biased.

6.3.3.2.2 Cross-validation error estimator

The cross-validation error estimator is a randomized estimator obtained by randomly selecting *K* folds. In the *K*-fold cross-validation error estimator, the data are split into

K folds at each step (*K*=10): one fold is used as a test ($S_m = \{(x_1, t_1), ..., (x_m, t_m)\}$), and the remaining folds are used for training ($S_{n'} = \{(x_1, t_1), ..., (x_{n'}, t_{n'})\}$), where m and n' are the number of samples in the test and training sets, respectively. The above procedure is repeated *K* times by leaving a different fold as test data, which are used to compute (estimate) the classification error. In each iteration, the estimated respective error is calculated as follows:

$$e_K = \frac{1}{m} \left\| (T_i - \phi_{n'}(x_i) \right\|_0 \tag{6.15}$$

Where $i = 1, ..., m$ .The total error is calculated using the average of the errors in each iteration.

$$e_{cross} = \frac{1}{K} \sum_{i=1}^{K} e_K \tag{6.16}$$

In this chapter, we used the standard *k*-nearest-neighbors (*k*-NN) estimator, with *k*=3, to compute the classification error estimation. The *k*-NN estimator is chosen due to its lower computational cost relative to that of a state-of-the-art SVM estimator.

6.3.3.3 Optimal number of features based on the classification error

In addition to the feature-ranking algorithm based on the discriminative performance of the features, we propose to use an automatic approach based on classification error estimation to determine the number of top discriminative features and, hence, reduce the dimensionality of prospective feature vectors. Using this approach, it is simpler to automatically determine the *q* top discriminative features based on the ranked values in the training data in each fold instead of using a fixed *q*. Once the features are ranked, the number of top ranked features iteratively increased from 1 to Γ (Γ <<*M*) in the respective training error estimation. *M* is the number of features in the respective feature vectors in each fold, which had values from 59395 to 69170 voxels

in our experiments. q is searched within the first Γ dimensions, where Γ is heuristically chosen to be 1500 to reduce the computational cost. Typical q values of between 10and 1300 are observed in Γ. q is regarded as the optimal number of top ranked features that minimizes the classification error in the training set. The proposed algorithm to determine q, is given in the pseudo code shown in algorithm 6-1. The number of top features is iteratively incremented from 1 to Γ, using a training set of each fold to calculate the respective classification error estimation values by the *k*-NN estimator. Using a cross-validation process, the optimal numbers of top features, q, minimizing the classification error estimation in training phase is selected for use as the optimal dimension in the test and the training data in each fold. Figure 6.2 shows the details of the proposed feature selection procedure.

Algorithm 6.1. Optimal number of top feature selection procedure based on the classification error.

1: $V \leftarrow component\_set(Data_{Train}, Label_{Train})$

2: $Ranked\ features \leftarrow feature\_ranking(Data_{Train}, Label_{Train})$

2: $number\ of\ top\ features \leftarrow \emptyset,\ \Gamma = 1500$

3: **for** $n = 1$ *to* Γ **do**

4: $\hat{e}(n) \leftarrow Ranked\ features(1:n, Label_{Train})$

5: **end for**

6: $q \leftarrow \arg\min_{n \in \{1,...,\Gamma\}} \hat{e}(n)$

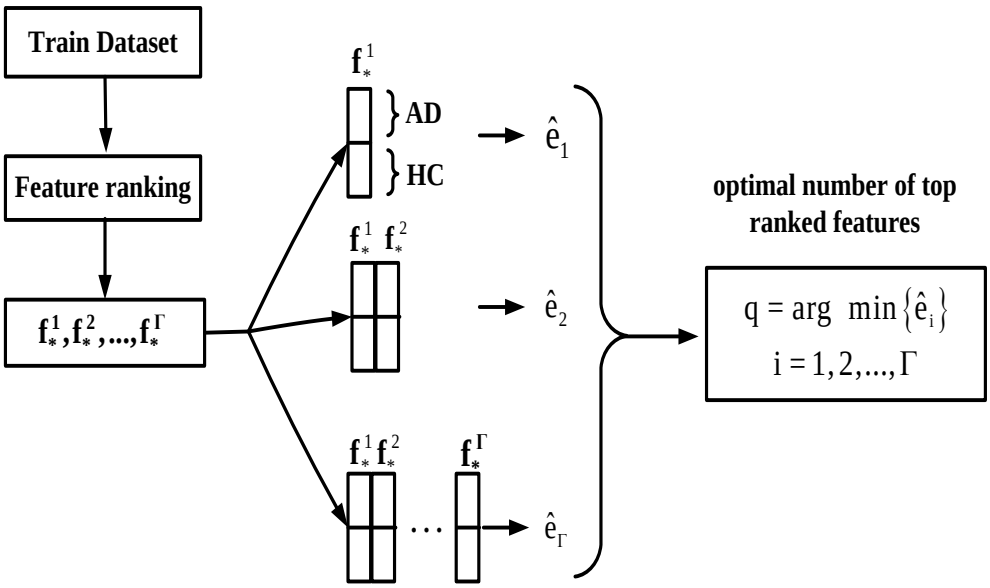


Figure 6.2: Detailed illustration of the proposed feature selection approach

6.3.4 Data fusion among different feature ranking methods

This chapter introduces a data fusion technique among different feature ranking methods to improve the performance of the proposed feature-ranking-based AD classification. The aim of the data fusion technique is to integrate the data from two or more distinct multiple sources to improve performance. The pipeline of the proposed data fusion system combining different feature ranking methods is illustrated in Figure 6.3. In the scheme of proposed data fusion, the top *q*-ranked features [f_*^1, f_*^2,..., f_*^q] selected based on approaches, described in section 6.3.3.1, from different feature ranking methods, are combined into a single feature vector using union operator. Assuming $FRV_1, FRV_2, ..., FRV_z$ are feature ranked vectors generated using different feature ranking methods. The feature vector fusion (FVF) is then:

$$FVF = \left[FRV_1 \cup FRV_2 \cup ... \cup FRV_z \right]_{1\times\theta} \tag{6.17}$$

where θ is the vector length for FVF, z is the number of ranked methods and $\Gamma \leq \theta$. This concatenated feature vector is then used for post-feature ranking. In this regard, the MI based feature ranking is used, because of its better performance in comparison to other ranking methods (see Table 6.2). The ranked feature vector fusion, followed by the determination of resubstitution and cross-validation error estimators to select the top features that minimizes the error in the ranked feature vector fusion set.

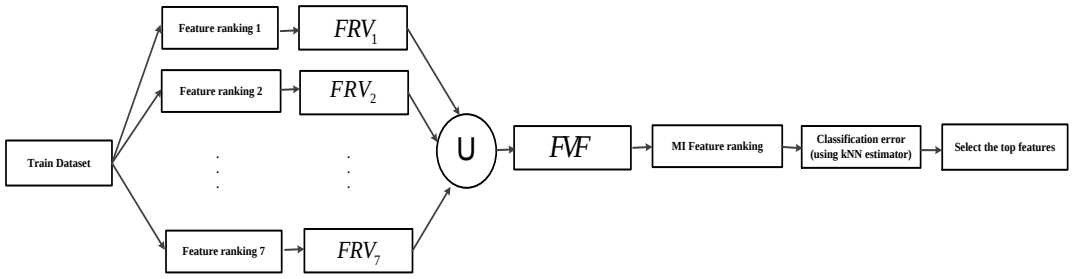


Figure 6.3: The pipeline of the proposed data fusion system combining different feature ranking methods

6.4 Experimental results and discussion

In this section, the experimental results obtained through the preprocessing phase using VBM plus DARTEL analysis on 3D T1weighted MR Imaging are considered, as an indicator disclosing significance of decreased gray matter volumes in ADs contributing to VOI. The experimental data consisted of 260 samples from an ADNI data set. A 10-fold cross-validation is employed throughout the performance analysis, with 234 (90%) samples in the training sample and 26 (10%) samples in the testing processes in each iteration. The performance of the classification is reported for the following cases: 1) The performance of raw feature vectors directly extracted from the VBM and 2) The performance of the proposed feature-ranking technique using the optimal number of top features based on the classification error. 3) The

performance of the proposed data fusion technique among different feature ranking methods. The ACC (%), SEN (%), SPE (%), and AUC performance metrics are used for the performance assessment. The AUC is a widely used measure of performance for classification and diagnostic rules(Hand, 2009).

6.4.1 VBM of GM analysis in AD versus HC

VBM plus DARTEL of GM analysis specified significant GM atrophy in the right/left hippocampus, right inferior parietal lobe, and right anterior cingulate in the ADs compared to the HCs. For an example, comparison of gray matter volume among 117 ADs and 117 HCs in fold 1 training is illustrated in Figure 6.4. The voxel locations of these significant regions are segmented as a 3D mask in each fold. This mask is employed to the gray matter density volume results from the segmentation step in the MRI data pre-processing to extract voxel values as raw feature vectors.

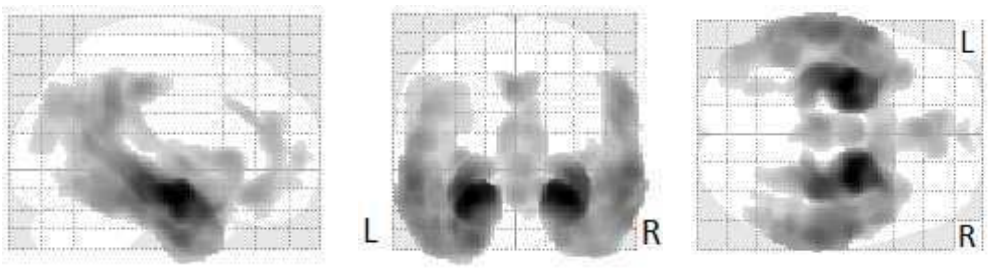


Figure 6.4: Brain regions with significant atrophy in gray matter volume in the 117 ADs compared to 117 HCs in fold 1

6.4.2 Performance of raw feature vectors

The complete MRI data set consisted of 130 AD and 130 HC samples. The ACC, SEN, SPE, and AUC obtained in the 10-fold cross validation using a linear SVM classifier on raw feature vectors are presented in Table 6.1.

Table 6.1: Raw feature vectors performance of atrophy clusters using 10 fold cross validation

Classifier	ACC(%)	SEN(%)	SPE(%)	AUC (%)
SVM-linear	83.58	82.04	85.12	92.10
SVM-RBF	86.02	89.70	82.35	93.13

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve.

6.4.3 Performance of the proposed feature-selection method using feature ranking and classification error

As introduced in Section 6.3.3.3, the proposed feature-selection techniques is evaluated by using seven different feature-ranking methods (SD, MI, IG, PCC, TS, FC, and GI),followed by two different classification errors(resubstitution and cross-validation error) to determine the optimal number of top features. Figure 6.5 shows the improvement in the ACC obtained by using progressive inclusion of the ranked features in the feature vector. A logarithmic scale is used to cover the entire feature space. This performance is reported for fold 1 after the MI feature ranking. The ACC is 80.76% and 92.30% on raw feature vectors and top 1500 ranked features after the MI feature ranking. The ACC performance improved with an increased number of ranked features, up to 96.15%. The performance level corresponded to the number of top ranked features, 479, which minimized the cross-validation error. The number of features that minimized the resubstitution error is 864, with an ACC performance of 92.30%. Table 6.2 shows the overall performances of the proposed feature-selection method. The results clearly show the performance improvement provided by the proposed feature-selection method. Among the seven different feature-ranking methods, in general, the MI generated the highest performance for both classification errors to determine the optimal size of the sample vectors. Regarding the

classification errors, the e_{cross} -based approach gives a higher performance than the e_{resub} -based method. The superior performance of the e_{cross} -based approach is attributed to the randomization in the cross validation, with the e_{cross} -based approach reducing the bias, which is the main problem of the e_{resub} -based method. Among the alternative methods tested, the results indicate that the MI feature ranking gives the highest or equal performance in terms of the ACC(%), SEN(%), SPE(%), and AUC(%), when compared with the other seven ranking methods. Recently, MI feature selection approach has been widely used for feature selection in pattern recognition studies (Z. Yan, Wang, & Xie, 2008; Yu & Lee, 2012).

Table 6.2: Performance results of the proposed feature selection method with linear SVM

	resubstitutionerror (e_{resub})				Cross validation error (e_{cross})			
	ACC(%)	SEN(%)	SPE(%)	AUC(%)	ACC(%)	SEN(%)	SPE(%)	AUC(%)
SD	86.92	83.07	90.76	94.38	89.61	88.46	90.76	95.74
MI	**88.84**	**86.92**	**90.76**	**94.20**	**91.53**	**90.00**	**93.07**	**95.80**
IG	88.07	87.69	88.46	94.93	88.07	86.92	89.23	94.50
PCC	86.15	86.92	85.38	94.97	89.23	91.53	86.92	94.62
TS	86.15	86.92	85.38	94.97	89.23	91.53	86.92	94.62
FC	86.15	86.92	85.38	94.97	89.23	91.53	86.92	94.62
GI	86.15	85.38	86.92	93.67	87.30	86.14	88.45	93.96

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; SD, statistical dependency; MI, mutual information; IG, information gain; PCC, Preason's correlation coefficient; TS, t-test score; FC, Fisher criterion; GI, Gini index.

Table 6.3: Performance results of the proposed feature selection method with nonlinear SVM

	resubstitutionerror (e_{resub})				Cross validation error (e_{cross})			
	ACC(%)	SEN(%)	SPE(%)	AUC(%)	ACC(%)	SEN(%)	SPE(%)	AUC(%)
SD	88.07	85.38	90.76	94.56	89.23	90.00	88.46	95.33
MI	**89.61**	**89.23**	**90.00**	**94.32**	**90.38**	**89.23**	**91.53**	**95.15**
IG	86.92	86.92	86.92	94.26	87.30	88.46	86.15	94.44
PCC	85.76	86.15	85.38	94.38	86.53	86.15	86.92	94.44
TS	85.76	86.15	85.38	94.38	86.53	86.15	86.92	94.44
FC	85.76	86.15	85.38	94.38	86.53	86.15	86.92	94.44
GI	85.38	83.84	86.92	93.73	86.15	84.61	87.69	92.78

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; SD, statistical dependency; MI, mutual information; IG, information gain; PCC, Preason's correlation coefficient; TS, t-test score; FC, Fisher criterion; GI, Gini index.

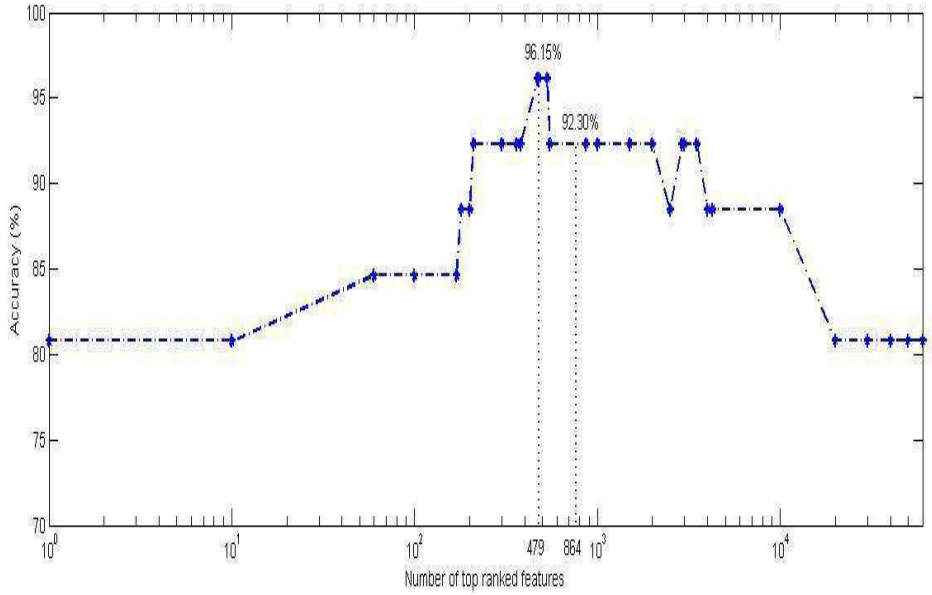


Figure 6.5: Accuracy (%) by different number of top ranked features selected using MI ranking in fold 1

6.4.4 Performance of proposed data fusion among different feature ranking methods

The performance improvement aided by proposed data fusion of seven different feature ranking methods is shown in Table 6.4. The performance of the e_{cross} -based approach is always higher than the performance of the e_{resub} -based method.

Table 6.4: Performance of proposed data fusion technique among feature ranking methods

resubstitutionerror (e_{resub})				Cross validation error (e_{cross})			
ACC(%)	**SEN(%)**	**SPE(%)**	**AUC(%)**	**ACC(%)**	**SEN(%)**	**SPE(%)**	**AUC(%)**
88.84	86.92	90.76	94.20	**92.48**	91.07	**93.89**	96.30

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve.

6.5 Discussion

This chapter investigated the feature ranking and classification errors as part of a novel feature-selection method to design an automatic CAD system for high-dimensional pattern classification in AD. In the proposed system, we evaluated seven feature ranking approaches to rank the features with respect to their statistical relevance. In addition, we proposed an automatic criterion to select the subset of top ranked features based on classification error in the training part. In this context, resubstitution and cross-validation error estimators are employed to identify the number of ranked features. By investigation Table 1 and Table 2, it is clear that proposed feature selection method significantly improved the performance with respect to raw feature vectors. For example, feature selection using MI ranking and cross-validation error estimator provided 8% improvement in accuracy in comparison to raw feature vectors. Many researchers studied Random Forest as alternative feature selection method in machine learning, because of its relatively good accuracy and robustness (Díaz-uriarte & Andrés, 2006; Ebina, Toh, & Kuroda,

2011; Genuer, Poggi, & Tuleau-malot, 2010). Otherwise, using Random Forest suffer from biased towards features with many categories and with correlated features, more informative features can end up with low scores (Strobl, Boulesteix, Zeileis, & Hothorn, 2007). In addition, several studies investigated high-dimensional pattern classification approach in a number of neuroimaging studies (I. Beheshti & Demirel, 2015b; Fan, Batmanghelich, et al., 2008; Fan et al., 2005; Lao et al., 2004). For example in (Fan, Batmanghelich, et al., 2008), the authors presented an advanced quantitative pattern analysis and classification of brain atrophy in MCI and AD patients. In (Fan et al., 2005) authors introduced a method based on Support Vector Machine-Recursive Feature Elimination (SVM-RFE) technique for feature ranking and they used SVM classifier for classification. Data used in the present study is the same as the one described in our previous study (I. Beheshti & Demirel, 2015b), including pre-processing steps and feature extraction. In (I. Beheshti & Demirel, 2015b), we introduced a novel statistical feature selection method based on the probability distribution function (PDF) of the VOI. In more detail, PDF is introduced to generate statistical pattern of the VOI representing the entire sMRI. Using proposed PDF-based method, we obtained 89.65% accuracy with linear SVM. In the present study, instead of generation of the statistical pattern of the VOI, we introduced an automatic statistical feature selection method based on the combination of feature ranking and the classification error of the VOI, which can be considered a lower-dimensional feature vector representation of sMRI. The dimensionality of the feature vector can be adjusted by minimizing the classification error in the training data-set. The proposed feature selection method not only selects the top discriminative features but also reduces the dimensionality of the input vectors to feature vectors. Finally, we proposed a data fusion technique among the

different feature ranking methods and obtained 91.97 % accuracy with linear SVM. As part of future studies on AD classification, we suggest considering feature ranking-based feature selection for high-dimensional pattern classification such as the deformation-based analysis. Another priority for future studies is to use other registration methods as described in(Klein et al., 2009). These methods could further be used to evaluate the accuracy of inter-subject registration in GM volume changes in patients with AD.

6.6 Performance comparison to other methods

Recently, several studies have reported classification results to distinguish AD patients and HCs based on MRI and ADNI dataset. Aguilar et al. (Aguilar et al., 2013) employed FreeSurfer software to compute cortical thickness and volumetric measures. Based on an artificial neural network classifier and MRI data, they achieved an ACC of 84.9% and an ACC of 88.8% using an SVM classifier and a combination of MRI data with educational and demographic data. Querbes et al. (Querbes et al., 2009) reported an ACC of 85%using a cortical thickness feature from MRI data. Khedher et al. (Khedher et al., 2015) achieved an ACC of 88.49% by combining GM and white matter modalities in MRI data. Cuingnet et al. (Cuingnet et al., 2011) tested 10 methods. They presented an SEN of 81% and an SPE of 95% as the best performances. Zhang et al.(D. Zhang et al., 2011) used a multimodal classification of AD based on a combination of MRI, CSF, and PET data. They reported an ACC of 86.2% in the classification of AD/HC using the MRI data. By combining the MRI, CSF, and PET results, they achieved a high ACC of 93.2%. Westman et al.(Westman et al., 2012) reported an ACC of 87% using MRI data and increased the ACC to 91.8% by combining the MRI data with CSF measures. Beheshti et al.(I. Beheshti & Demirel, 2015b) employed a PDF-based approach using

MRI data and reported an ACC of 89.65%. A comparison of the classification performance using the different methods and MRI data is provided in Table 6.5. The results show that the performance of the proposed feature-selection method using only MRI data is higher or comparable to that of other methods reported in the literature.

Table 6.5: Supervised classification results of Alzheimer's disease and healthy control subjects based on MRI from ADNI data-set

Author	AD/HC	Validation method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
Aguilar et al.2013 (Aguilar et al., 2013)	116/110	10 Fold	84.90	80.20	90.00	88.00
Querbes et al.,2009 (Querbes et al., 2009)	130/130	10 Fold	85.00	-	-	-
Khedher et al., 2015 (Khedher et al., 2015)	188/229	10 Fold	88.49	85.11	91.27	-
Cuingnet et al., 2011(Cuingnet et al., 2011)*	162/137	2 Fold	-	81.00	95.00	-
Zhang et al.,2011(D. Zhang et al., 2011)	51/52	10 Fold	86.20	86.00	86.30	-
Westman et al., 2012 (Westman et al., 2012)	96/111	10 Fold	87.00	83.30	90.10	93.00
Beheshti et al.,2015(I. Beheshti & Demirel, 2015b)	130/130	10 Fold	89.65	87.73	91.57	95.30
Proposed method	**130/130**	**10 Fold**	**92.48**	**91.07**	**93.89**	**93.30**

*This paper compares ten methods and the best performance is presented here.

6.7 Conclusion

This chapter proposed an automatic CAD system for the classification of AD based on seven feature-ranking methods (i.e., SD, MI, IG, PCC, TS, FC, and GI) and classification errors (i.e., resubstitution and cross-validation errors). The optimal size of the selected features is determined by classification error estimation, which minimized the classification error in the training phase. This approach is applied to extracted raw features obtained from GM atrophy clusters of VOIs, which are determined using a VBM analysis. An SVM classifier is used for the classification of the extracted feature vectors after the feature selection. A performance improvement is also proposed by applying data fusion among the different feature ranking based. The performance of the proposed system is evaluated with 10-fold cross validation

using an ADNI data set made up of 260 subjects (130 AD patients and 130 HCs). The results clearly showed that the proposed feature-selection method is a reliable technique for high-dimensional data. The experimental results showed that the performance of the proposed approach using only MRI data is higher or comparable to that of alternative methods reported in the literature.

Chapter 7

COMPARISON OF PROPOSED METHODS

7.1 Introduction

The presented thesis introduces three main methods feature selection approaches for high-dimensional classification of AD and HC. MRI biomarker is used for feature extraction, selection and classification. In the current study, the feature selection is applied to overall atrophy clusters determined by using VBM analysis. This procedure helps to select some regions of brain to reveal significant differences between ADs and HCs and select most discriminative features from brain. The current thesis has investigated several advanced feature selection approaches aimed at the high accurate identification of AD and HC. In Chapter 4, an automatic statistical feature-selection method, namely, PDF, is proposed for the classification of AD which can be considered a lower-dimensional feature vector representation of sMRI images. The proposed feature-selection method compresses the statistical information of high-dimensional data into a lower-dimensional vector. This approach is used for high-dimensional classification, especially for feature-extracted VOIs of gray matter atrophy. In addition, an automatic approach based on the Fisher criterion is introduced to determine the optimal number of bins of the histogram generating the PDF. This approach adaptively determines the number of PDF bins based on the training data in each fold instead of using a fixed one. The proposed PDF-based feature-selection method is evaluated using 130 AD and 130 HC MRI data with 10-

fold cross validation. The experimental results using the proposed PDF-based approach with SVM by linear Kernel generates 89.65% accuracy, 87.73% sensitivity, 91.57% specificity, and 95.33% AUC. In the chapter 5, a novel feature selection approach based on t-test feature ranking and Fisher Criterion is proposed for high-dimensional pattern recognition in AD detection. In the proposed approach, the number of top features is determined by using Fisher Criterion, which maximizes the class separation between AD and HC. In addition, data fusion techniques among different gray matter atrophy clusters in the brain are introduced to improve the classification performance. The performance of proposed system on 136 subjects (including 68 AD and 68 HC) is investigated using 10 fold cross validation. The proposed method yields accuracy (96.32%), sensitivity (94.11%), specificity (98.52%) and AUC (99.93%) for AD classification. finally, in the Chapter 6, an automatic and novel feature selection approach based on different feature ranking and classification error is investigated. In this regard, seven feature-ranking methods, namely, SD, MI, IG, PCC, TS, FC, and GI are evaluated in proposed feature selection method. Regarding to stopping criteria of the increasing dimensionality among the ranked features, the resubstitution and cross-validation error estimators are employed to estimate the optimal number of features with the most discriminative information in the classification process. In addition, a data fusion technique is proposed to improve the AD classification performance among different feature ranking methods. The performance of the proposed system is evaluated using a data set containing 260 subjects (130 AD patients and 130 HCs) using 10-fold cross-validation. The experimental results generate accuracy (92.48%), sensitivity (91.07%), specificity (93.89%), and area under the curve (96.30%) of the proposed method. Table 7.1 presents a comparison of the classification results based on

proposed approaches for high dimensional pattern recognition in AD detection. In order to provide a fair comparison, all methods are reevaluated using common dataset described in chapter 4, containing 260 subjects (130 AD patients and 130 HCs) using 10-fold cross-validation strategy.

Table 7.1: Comparison of classification performance from Chapters 4,5 and 6 with linear SVM

	Method	Stopping criteria	ACC(%)	SEN(%)	SPE(%)	AUC(%)
	PDF based	FC	89.65	87.73	91.57	95.33
		e_{resub}	88.46	86.15	90.76	95.95
		e_{cross}	89.23	86.92	91.53	96.57
Ranking based	SD	FC	87.30	86.92	87.69	95.38
		e_{resub}	86.92	83.07	90.76	94.38
		e_{cross}	89.61	88.46	90.76	95.74
	MI	FC	88.07	86.92	89.23	95.27
		e_{resub}	88.84	86.92	90.76	94.20
		e_{cross}	**91.53**	90.00	**93.07**	**95.80**
	IG	FC	86.53	83.07	90.00	93.67
		e_{resub}	88.07	87.69	88.46	94.93
		e_{cross}	88.07	86.92	89.23	94.50
	PCC	FC	86.92	85.38	88.46	94.08
		e_{resub}	86.15	86.92	85.38	94.97
		e_{cross}	89.23	**91.53**	86.92	94.62
	TS	FC	86.92	85.38	88.46	94.08
		e_{resub}	86.15	86.92	85.38	94.97
		e_{cross}	89.23	**91.53**	86.92	94.62
	FC	FC	86.92	85.38	88.46	94.08
		e_{resub}	86.15	86.92	85.38	94.97
		e_{cross}	89.23	**91.53**	86.92	94.62
	GI	FC	86.92	85.38	88.46	94.08
		e_{resub}	86.15	85.38	86.92	93.67
		e_{cross}	87.30	86.14	88.45	93.96

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; SD, statistical dependency; MI, mutual information; IG, information gain; PCC, Preason's correlation coefficient; TS, t-test score; FC, Fisher criterion; GI, Gini index; PDF, probability distribution function.

Regarding the accuracy, the MI-based ranking followed by e_{cross} shows the highest result in comparison to the other methods (90.53%). According to sensitivity, the maximum is achieved with PCC, TS, FC which are categorized in ranking-based methods followed by e_{cross} (91.53%). Based on specificity, the higher achievement is obtained with MI-based ranking followed by e_{cross} (93.07%) and regarding the AUC, the maximum is attained using MI feature ranking followed by e_{cross} (95.70%). In PDF-based feature selection, selecting the number of PDF bins using FC shows higher performance in ACC, SEN and SPE in comparison to the other stopping criteria. The only exception is AUC, where the performance using e_{cross} is higher than FC. Among the adapted stopping criteria to select the optimum number of top features, e_{cross} shows better performance in comparison to the other stopping criteria such as e_{resub} and FC among the alternative ranking methods. In determining the stopping criteria, error estimation, not only the most discriminative features are selected, but also the classification error in training phase is minimized. Minimizing this error corresponds to maximizing training accuracy and learning in the classification process. As an example, Table 7.2 shows the training accuracy based on MI-based feature ranking and three different stopping criteria using linear SVM. As shown in Table 7.2, it is clear that training accuracy of the e_{cross}-based approach is higher than e_{resub} and FC due to the randomization in the cross validation which helps to reduce the bias, which is the main problem of the e_{resub}-based method. Additionally, PCC, FC and TS ranking based approaches show the similar performances among the alternative stopping methods.

Table 7.2: Training accuracy base on MI feature ranking and three different stopping criteria

Stopping Criteria	ACC(%)	SEN(%)	SPE(%)	AUC(%)
FC	90.29	88.71	91.88	95.70
e_{cross}	**92.43**	**91.53**	93.33	**96.80**
e_{resub}	92.05	90.68	**93.41**	96.20

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve.

Table 7.3 shows the performance improvement by the help of proposed data fusion among seven different ranking methods as described in section 6.3.4. The performance of the e_{cross} -based approach is always higher than the performance of the FC and e_{resub} -based method.

Table 7.3: Performance of data fusion technique among feature ranking methods

Stopping Criteria	ACC(%)	SEN(%)	SPE(%)	AUC
FC	87.30	86.92	87.69	95.38
e_{cross}	**92.48**	**91.07**	**93.89**	**96.30**
e_{resub}	88.84	86.92	90.76	94.20

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve.

Chapter 8

CONCLUSION AND FUTURE WORK

8.1 Conclusion

In summary, the aim of this study is to introduce novel and automatic statistical feature selection methods for high-dimensional pattern recognition for AD detection. In Chapter 4, an automatic CAD technique is introduced based on a statistical feature-selection process, namely, PDF of VOI, for the classification of AD. The proposed feature-selection method compresses the statistical information of high-dimensional data into a lower-dimensional PDF vector. This approach is used for high-dimensional classification, especially for VOI of gray matter atrophy. The PDF-based feature-selection approach is compared to the standard PLS-based classification using SVM classifiers. Chapter 4 demonstrated that the PDF-based feature-selection method is a reliable alternative to the PLS-based method. The proposed PDF-based method not only extracts the selected statistical features but also reduces the dimensionality of the input vectors to feature vectors with acceptably low dimensions. In addition, dimensionality is determined using changing bin size based on the Fisher criterion to determine the optimal number of bins of the histogram generating the PDF. The optimal number of bins is obtained by maximizing the Fisher criterion among the possible number of bins.

Chapter 5 presented a feature selection method using t-test based feature ranking which is used for the classification of AD. The optimal size of the selected features is

determined by using Fisher Criterion, which maximizes the class separation between AD and HC. A performance improvement is also proposed by applying data fusion among the individual atrophy clusters as well as the overall atrophy cluster. In Chapter 6, a CAD system for the classification of AD based on feature-ranking method and classification errors is proposed. In this regard, seven-feature ranking method (i.e., SD, MI, IG, PCC, TS, FC, and GI) are evaluated. The optimal size of the selected features is determined by the classification error estimation, which minimizes the classification error in the training phase. Among the alternative methods tested, the results indicate that the MI feature ranking gives the highest or equal performance, when compared with the other seven ranking methods. In addition, a data fusion approach among feature ranking methods is introduced to improve the classification performance. Finally, Chapter 7 provided a comparison based on proposed approaches for high dimensional pattern recognition in AD detection. In summary, the results indicate that the MI feature ranking gives the highest performance, when compared with the other methods. The optimal size of the selected features is determined using the FC and classification error as stopping criteria. In this regard, e_{cross} shows a superior performance compared to the e_{resub} and FC.

8.2 Future work

As part of future studies on AD classification, we suggest considering feature ranking-based and PDF-based feature selection for high-dimensional pattern classification such as the deformation-based analysis and diffusion tensor imaging (Stebbins & Murphy, 2010; Teipel et al., 2007). Another priority for future studies is to use other registration methods such as (Klein et al., 2009). These methods could further be used to evaluate the accuracy of inter-subject registration in GM volume

changes in patients with AD. In addition, it is suggested to use data fusion techniques for the proposed MRI modality with other modalities, such as PET, CSF, and WM, and to combine them using the proposed approaches in order to achieve higher accuracy. For example, the PDF-based data fusion technique has already been used successfully in recent studies for the improvement of face-recognition performance (H Demirel & Anbarjafari, n.d.; Hasan Demirel & Anbarjafari, 2009). One interesting area for further research could be to use the heuristic methods such as genetic algorithm to select an optimal feature subset. Another priority for future studies is to employ 3D wavelet analysis in pre-processing stage and use data fusion techniques among different sub-bands to increase the performance.

REFERENCES

3D Brain Image-based Diagnosis of Alzheimer's Disease : Bringing Medical Vision into Feature Selection. (2012), 134–137.

Aguilar, C., Westman, E., Muehlboeck, J. S., Mecocci, P., Vellas, B., Tsolaki, M., … Wahlund, L. O. (2013). Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. *Psychiatry Research - Neuroimaging*, *212*(2), 89–98. http://doi.org/10.1016/j.pscychresns.2012.11.005

Al-Kadi, O. S. (2014). A multiresolution clinical decision support system based on fractal model design for classification of histological brain tumours. *Computerized Medical Imaging and Graphics*, *41*, 67–79. http://doi.org/10.1016/j.compmedimag.2014.05.013

Alternative Feature Extraction Methods In 3D Brain Image-Based Diagnosis Of Alzheimer's Disease. (2012), (1), 1237–1240.

Alzheimer's Association | Alzheimer's Disease and Dementia. (2015). Retrieved April 5, 2015, from http://www.alz.org/

Andersen, A. H., Rayens, W. S., Liu, Y., & Smith, C. D. (2012). Partial least squares for discrimination in fMRI data. *Magnetic Resonance Imaging*, *30*(3), 446–452. http://doi.org/10.1016/j.mri.2011.11.001

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*(1), 95–113. http://doi.org/10.1016/j.neuroimage.2007.07.007

Ashburner, J. (2009). Computational anatomy with the SPM software. *Magnetic Resonance Imaging*, *27*(8), 1163–1174. http://doi.org/10.1016/j.mri.2009.01.006

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry--the methods. *NeuroImage*, *11*(6 Pt 1), 805–821. http://doi.org/10.1006/nimg.2000.0582

Baron, J. C., Chételat, G., Desgranges, B., Perchey, G., Landeau, B., de la Sayette, V., & Eustache, F. (2001). In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. *NeuroImage*, *14*(2), 298–309. http://doi.org/10.1006/nimg.2001.0848

Beg, M. F., Raamana, P. R., Barbieri, S., & Wang, L. (2012). Comparison of four shape features for detecting hippocampal shape changes in early Alzheimer's. *Statistical Methods in Medical Research*. http://doi.org/10.1177/0962280212448975

Beheshti, I., & Demirel, H. (2015a). Feature-ranking-based Alzheimer's disease classification from structural MRI. *Magnetic Resonance Imaging*. http://doi.org/10.1016/j.mri.2015.11.009

Beheshti, I., & Demirel, H. (2015b). Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease.

Computers in Biology and Medicine, *64*, 208–216. http://doi.org/10.1016/j.compbiomed.2015.07.006

Beheshti, I., Demirel, H., & Yang, C. (2015). Significance of Sex Differences on Gray Matter Atrophy in Alzheimer ' s Disease : A Voxel-Based Morphometry Study. *British Biomedical Bulletin*, *3*(4), 522–536.

Ben Ahmed, O., Benois-Pineau, J., Allard, M., Ben Amar, C., & Catheline, G. (2014). Classification of Alzheimer's disease subjects from MRI using hippocampal visual features. *Multimedia Tools and Applications*, *74*(4), 1249–1266. http://doi.org/10.1007/s11042-014-2123-y

Braga-Neto, U. M. (2009). Classification and error estimation for discrete data. *Current Genomics*, *10*(7), 446–462. http://doi.org/10.2174/138920209789208228

Bron, E. E., Smits, M., van der Flier, W. M., Vrenken, H., Barkhof, F., Scheltens, P., … Klein, S. (2015). Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge. *NeuroImage*, *111*, 562–579. http://doi.org/10.1016/j.neuroimage.2015.01.048

Cabral, C., Morgado, P. M., Campos Costa, D., & Silveira, M. (2015). Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages. *Computers in Biology and Medicine*, *58*, 101–109. http://doi.org/10.1016/j.compbiomed.2015.01.003

Carter, C. L., Resnick, E. M., Mallampalli, M., & Kalbarczyk, A. (2012). Sex and Gender Differences in Alzheimer's Disease: Recommendations for Future Research. *Journal of Women's Health*. http://doi.org/10.1089/jwh.2012.3789

Casanova, R., Maldjian, J. a., & Espeland, M. a. (2011). Evaluating the Impact of Different Factors on Voxel-Based Classification Methods of ADNI Structural MRI Brain Images. *International Journal of Biomedical Data Mining*, *1*, 1–10. http://doi.org/10.4303/ijbdm/B110102

Challis, E., Hurley, P., Serra, L., Bozzali, M., Oliver, S., & Cercignani, M. (2015). Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. *NeuroImage*, *112*, 232–243. http://doi.org/10.1016/j.neuroimage.2015.02.037

Chang, Y.-W., & Lin, C.-J. (2008). Feature ranking using linear svm. *JMLR:Workshop and Confrence Proceedings*, *3*, 53–64.

Chaves, R., Ram, J., Segovia, F., & Padilla, P. (2009). SPECT image classification based on NMSE feature correlation weighting and SVM, *1*(3).

Chaves, R., Ramírez, J., Górriz, J. M., López, M., Salas-Gonzalez, D., Álvarez, I., & Segovia, F. (2009). 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. http://doi.org/10.1016/j.neulet.2009.06.052

Chaves, R., Ramírez, J., Górriz, J. M., & Puntonet, C. G. (2012). Association rule-based feature selection method for Alzheimer's disease diagnosis. *Expert Systems with Applications*, *39*(14), 11766–11774. http://doi.org/10.1016/j.eswa.2012.04.075

Chen, Y. J., Deutsch, G., Satya, R., Liu, H. G., & Mountz, J. M. (2013). A semi-quantitative method for correlating brain disease groups with normal controls using SPECT: Alzheimer's disease versus vascular dementia. *Computerized Medical Imaging and Graphics*, *37*(1), 40–47. http://doi.org/10.1016/j.compmedimag.2012.11.001

Chételat, G., Landeau, B., Eustache, F., Mézenge, F., Viader, F., De La Sayette, V., … Baron, J. C. (2005). Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *NeuroImage*, *27*(4), 934–946. http://doi.org/10.1016/j.neuroimage.2005.05.015

Chincarini, A., Bosco, P., Calvini, P., Gemme, G., Esposito, M., Olivieri, C., … Nobili, F. (2011). Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *NeuroImage*, *58*(2), 469–480. http://doi.org/10.1016/j.neuroimage.2011.05.083

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., … Colliot, O. (2009). Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus*, *19*(6), 579–587.

http://doi.org/10.1002/hipo.20626

Coupé, P., Eskildsen, S. F., Manjón, J. V., Fonov, V. S., & Collins, D. L. (2012). Simultaneous segmentation and grading of anatomical structures for patient's classification: Application to Alzheimer's disease. *NeuroImage*, *59*(4), 3736–3747. http://doi.org/10.1016/j.neuroimage.2011.10.080

Cousijn, J., Wiers, R. W., Ridderinkhof, K. R., Van den Brink, W., Veltman, D. J., & Goudriaan, A. E. (2012). Grey matter alterations associated with cannabis use: Results of a VBM study in heavy cannabis users and healthy controls. *NeuroImage*, *59*(4), 3845–3851. http://doi.org/10.1016/j.neuroimage.2011.09.046

Cuadra, M. B., Cammoun, L., Butz, T., Cuisenaire, O., & Thiran, J. P. (2005). Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. *IEEE Transactions on Medical Imaging*, *24*(12), 1548–1565. http://doi.org/10.1109/TMI.2005.857652

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M. O., … Colliot, O. (2011). 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. http://doi.org/10.1016/j.neuroimage.2010.06.013

Demirel, H., & Anbarjafari, G. (n.d.). Improved PDF Based Face Recognition Using Data Fusion. *Journal of Electronics; Mechanical and Mechatronics Engineering*, *2*(2), 195–200. Retrieved from

http://www.aydin.edu.tr/ijemme/articles/vol2num2/improved_pdf_based_face_recognition_using_data_fusion.pdf

Demirel, H., & Anbarjafari, G. (2008). Pose Invariant Face Recognition Using Probability Distribution Functions in Different Color Channels. *Signal Processing Letters*, *15*, 537 – 540. http://doi.org/10.1109/LSP.2008.926729

Demirel, H., & Anbarjafari, G. (2009). Data fusion boosted face recognition based on probability distribution functions in different colour channels. *Eurasip Journal on Advances in Signal Processing*, *2009*. http://doi.org/10.1155/2009/482585

Diaf, a., Boufama, B., & Benlamri, R. (2013). Non-parametric Fisher's discriminant analysis with kernels for data classification. *Pattern Recognition Letters*, *34*(5), 552–558. http://doi.org/10.1016/j.patrec.2012.10.030

Díaz-uriarte, R., & Andrés, S. A. De. (2006). Gene selection and classification of microarray data using random forest, *13*, 1–13. http://doi.org/10.1186/1471-2105-7-3

Dimitrovski, I., Kocev, D., Kitanovski, I., Loskovska, S., & Džeroski, S. (2015). Improved medical image modality classification using a combination of visual and textual features. *Computerized Medical Imaging and Graphics*, *39*, 14–26. http://doi.org/10.1016/j.compmedimag.2014.06.005

Dinesh, E., Kumar, M. S., Vigneshwar, M., & Mohanraj, T. (2013). Instinctive

classification of Alzheimer's disease using FMRI, pet and SPECT images. *2013 7th International Conference on Intelligent Systems and Control (ISCO)*, 405–409. http://doi.org/10.1109/ISCO.2013.6481189

Duch, W., Wieczorek, T., Biesiada, J., & Blachnik, M. (2004). Comparison of feature ranking methods based on information entropy. *IEEE International Conference on Neural Networks - Conference Proceedings*, *2*, 1415–1419. http://doi.org/10.1109/IJCNN.2004.1380157

Duchesne, S., Caroli, A., Geroldi, C., Barillot, C., Frisoni, G. B., & Collins, D. L. (2008). MRI-based automated computer classification of probable AD versus normal controls. *IEEE Transactions on Medical Imaging*, *27*(4), 509–520. http://doi.org/10.1109/TMI.2007.908685

Dukart, J., Mueller, K., Barthel, H., Villringer, A., Sabri, O., & Schroeter, M. L. (2013). Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. *Psychiatry Research - Neuroimaging*, *212*(3), 230–236. http://doi.org/10.1016/j.pscychresns.2012.04.007

Ebina, T., Toh, H., & Kuroda, Y. (2011). DROP : an SVM domain linker predictor trained with optimal features selected by random forest, *27*(4), 487–494. http://doi.org/10.1093/bioinformatics/btq700

Fan, Y., Batmanghelich, N., Clark, C. M., & Davatzikos, C. (2008). Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern

classification, predict subsequent cognitive decline. *NeuroImage*, *39*(4), 1731–43. http://doi.org/10.1016/j.neuroimage.2007.10.031

Fan, Y., Resnick, S. M., Wu, X., & Davatzikos, C. (2008). Structural and functional biomarkers of prodromal Alzheimer's disease: A high-dimensional pattern classification study. *NeuroImage*, *41*(2), 277–285. http://doi.org/10.1016/j.neuroimage.2008.02.043

Fan, Y., Shen, D., & Davatzikos, C. (2005). Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. *Medical Image Computing and Computer-Assisted Intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, *8*(Pt 1), 1–8. http://doi.org/10.1007/11566465_1

Friston, K. J. (2006). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. *Functional neuroimaging: Technical*.

Fung, G., & Stoeckel, J. (2007). SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. *Knowledge and Information Systems*, *11*(2), 243–258. http://doi.org/10.1007/s10115-006-0043-5

Gao, Q., Liu, J., Zhang, H., Hou, J., & Yang, X. (2012). Enhanced fisher discriminant criterion for image recognition. *Pattern Recognition*, *45*(10), 3717–3724. http://doi.org/10.1016/j.patcog.2012.03.024

Geng, X., Liu, T., Qin, T., & Li, H. (2007). Feature Selection for Ranking. *Sigir*, (49), 407–414.

Genuer, R., Poggi, J., & Tuleau-malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225–2236. http://doi.org/10.1016/j.patrec.2010.03.014

Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H. S., … Colliot, O. (2009). Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage*, *47*(4), 1476–1486. http://doi.org/10.1016/j.neuroimage.2009.05.036

Górriz, J. M., Segovia, F., Ramírez, J., Lassl, a., & Salas-Gonzalez, D. (2011). GMM based SPECT image classification for the diagnosis of Alzheimer's disease. *Applied Soft Computing*, *11*(2), 2313–2325. http://doi.org/10.1016/j.asoc.2010.08.012

Graña, M., Termenon, M., Savio, a., Gonzalez-Pinto, a., Echeveste, J., Pérez, J. M., & Besga, a. (2011). Computer Aided Diagnosis system for Alzheimer Disease using brain Diffusion Tensor Imaging features selected by Pearson's correlation. *Neuroscience Letters*, *502*(3), 225–229. http://doi.org/10.1016/j.neulet.2011.07.049

Gray, K. R., Wolz, R., Heckemann, R. a., Aljabar, P., Hammers, A., & Rueckert, D. (2012). Multi-region analysis of longitudinal FDG-PET for the classification of

Alzheimer's disease. *NeuroImage*, *60*(1), 221–229. http://doi.org/10.1016/j.neuroimage.2011.12.071

Gray, K. R., Wolz, R., Keihaninejad, S., & Heckemann, R. A. (2011). Regional Analysis of FDG-PET for use in The Classification of Alzheimer's Disease, 1082–1085.

Greve, D. N. (2011). An Absolute Beginner ' s Guide to Surface- and Voxel-based Morphometric Analysis. *Presented at ISMRM 19th Annual Meeting & Exhibition; Montréal, Québec, Canada*, 7–13.

Guo, X., Wang, Z., Li, K., Li, Z., Qi, Z., Jin, Z., … Chen, K. (2010). Voxel-based assessment of gray and white matter volumes in Alzheimer's disease. *Neuroscience Letters*, *468*(2), 146–150. http://doi.org/10.1016/j.neulet.2009.10.086

Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, *77*(1), 103–123. http://doi.org/10.1007/s10994-009-5119-5

Hanyu, H., Sato, T., Hirao, K., Kanetaka, H., Iwamoto, T., & Koizumi, K. (2010). The progression of cognitive deterioration and regional cerebral blood flow patterns in Alzheimer's disease: A longitudinal SPECT study. *Journal of the Neurological Sciences*, *290*(1-2), 96–101. http://doi.org/10.1016/j.jns.2009.10.022

Haq, N. F., Kozlowski, P., Jones, E. C., Chang, S. D., Goldenberg, S. L., & Moradi, M. (2015). A data-driven approach to prostate cancer detection from dynamic contrast enhanced MRI. *Computerized Medical Imaging and Graphics*, *41*, 37–45. http://doi.org/10.1016/j.compmedimag.2014.06.017

Heijden, V. F. Van Der, & Ridder, D. De. (2004). *Classification, parameter estimation, and state estimation: an engineering Approach using MATLAB*. *Journal of Time Series Analysis* (Vol. 32). John Wiley and Sons.

Hinrichs, C., Singh, V., Xu, G., & Johnson, S. (2009). MKL for robust multi-modality AD classification. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 5762 LNCS, pp. 786–794). http://doi.org/10.1007/978-3-642-04271-3_95

Hinrichs, C., Singh, V., Xu, G., & Johnson, S. C. (2011). Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *NeuroImage*, *55*(2), 574–589. http://doi.org/10.1016/j.neuroimage.2010.10.081

Hirata, Y., Matsuda, H., Nemoto, K., Ohnishi, T., Hirao, K., Yamashita, F., … Samejima, H. (2005). Voxel-based morphometry to discriminate early Alzheimer's disease from controls. *Neuroscience Letters*, *382*(3), 269–274. http://doi.org/10.1016/j.neulet.2005.03.038

Huang, C. H. C., Yan, B. Y. Bin, Jiang, H. J. H., & Wang, D. W. D. (2008).

Combining Voxel-based Morphometry with Artifical Neural Network Theory in the Application Research of Diagnosing Alzheimer's Disease. *2008 International Conference on BioMedical Engineering and Informatics*, *1*, 250–254. http://doi.org/10.1109/BMEI.2008.245

Illán, I. A., Górriz, J. M., Ramírez, J., Salas-gonzalez, D., López, M. M., Segovia, F., & Chaves, R. (2011). F-FDG PET imaging analysis for computer aided Alzheimer ' s diagnosis q, *181*, 903–916. http://doi.org/10.1016/j.ins.2010.10.027

Kamkar, I., Gupta, S. K., Phung, D., & Venkatesh, S. (2014). Stable feature selection for clinical prediction: Exploiting ICD tree structure using Tree-Lasso. *Journal of Biomedical Informatics*, *53*, 277–290. http://doi.org/10.1016/j.jbi.2014.11.013

Karas, G. B., Burton, E. J., Rombouts, S. a R. B., Van Schijndel, R. a., O'Brien, J. T., Scheltens, P., … Barkhof, F. (2003). A comprehensive study of gray matter loss in patients with Alzheimer's disease using optimized voxel-based morphometry. *NeuroImage*, *18*(4), 895–907. http://doi.org/10.1016/S1053-8119(03)00041-7

Kasahara, K., Hashimoto, K., Abo, M., & Senoo, A. (2012). Voxel- and atlas-based analysis of diffusion tensor imaging may reveal focal axonal injuries in mild traumatic brain injury - comparison with diffuse axonal injury. *Magnetic Resonance Imaging*, *30*(4), 496–505. http://doi.org/10.1016/j.mri.2011.12.018

Khedher, L., Ramírez, J., Górriz, J. M., Brahim, a., & Segovia, F. (2015). Early diagnosis of Alzheimer′s disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing*, *151*, 139–150. http://doi.org/10.1016/j.neucom.2014.09.072

Kim, J., & Lee, J.-H. (2013). Integration of structural and functional magnetic resonance imaging improves mild cognitive impairment detection. *Magnetic Resonance Imaging*, *31*(5), 718–32. http://doi.org/10.1016/j.mri.2012.11.009

Klein, A., Andersson, J., Ardekani, B. a., Ashburner, J., Avants, B., Chiang, M. C., … Parsey, R. V. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, *46*(3), 786–802. http://doi.org/10.1016/j.neuroimage.2008.12.037

Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., … Frackowiak, R. S. J. (2008). Automatic classification of MR scans in Alzheimer's disease. *Brain*, *131*(3), 681–689. http://doi.org/10.1093/brain/awm319

Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S. M., & Davatzikos, C. (2004). Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage*, *21*(1), 46–57. http://doi.org/10.1016/j.neuroimage.2003.09.027

Lee, W., Park, B., & Han, K. (2013). Classification of diffusion tensor images for the early detection of Alzheimer's disease. *Computers in Biology and Medicine*,

43(10), 1313–1320. http://doi.org/10.1016/j.compbiomed.2013.07.004

Li, M., Qin, Y., Gao, F., Zhu, W., & He, X. (2014). Discriminative analysis of multivariate features from structural MRI and diffusion tensor images. *Magnetic Resonance Imaging*, *32*(8), 1043–1051. http://doi.org/10.1016/j.mri.2014.05.008

Li, S., Shi, F., Pu, F., Li, X., Jiang, T., Xie, S., & Wang, Y. (2007). Hippocampal shape analysis of Alzheimer disease based on machine learning methods. *American Journal of Neuroradiology*, *28*(7), 1339–1345. http://doi.org/10.3174/ajnr.A0620

Liu, M., Zhang, D., & Shen, D. (2012). Ensemble sparse classification of Alzheimer's disease. *NeuroImage*, *60*(2), 1106–1116. http://doi.org/10.1016/j.neuroimage.2012.01.055

Liu, S., Cai, W., Wen, L., Feng, D. D., Pujol, S., Kikinis, R., … Eberl, S. (2014). Multi-Channel neurodegenerative pattern analysis and its application in Alzheimer's disease characterization. *Computerized Medical Imaging and Graphics*, *38*(6), 436–444. http://doi.org/10.1016/j.compmedimag.2014.05.003

Magnin, B., Mesrob, L., & Kinkingnéhun, S. (2009). Support vector machine-based classification of Alzheimer ' s disease from whole-brain anatomical MRI, 73–83. http://doi.org/10.1007/s00234-008-0463-x

Manjón, J. V., Coupé, P., Martí-Bonmatí, L., Collins, D. L., & Robles, M. (2010).

Adaptive non-local means denoising of MR images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging*, *31*(1), 192–203. http://doi.org/10.1002/jmri.22003

Matsuda, H., Mizumura, S., Nemoto, K., Yamashita, F., Imabayashi, E., Sato, N., & Asada, T. (2012). Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated lie algebra improves the diagnosis of probable Alzheimer disease. *American Journal of Neuroradiology*, *33*(6), 1109–1114. http://doi.org/10.3174/ajnr.A2935

Mesrob, L. (2012). DTI and Structural MRI Classification in Alzheimer's Disease. *Advances in Molecular Imaging*, *02*(02), 12–20. http://doi.org/10.4236/ami.2012.22003

Mikhno, A., Nuevo, P. M., Devanand, D. P., Parsey, R. V, & Laine, A. F. (2012). Multimodal Classification of Dementia using Functional Data , Anatomical Features and 3D Invariant Shape Descriptors, 606–609.

Modi, S., Bhattacharya, M., Singh, N., Tripathi, R. P., & Khushu, S. (2012). Effect of visual experience on structural organization of the human brain: A voxel based morphometric study using DARTEL. *European Journal of Radiology*, *81*(10), 2811–2819. http://doi.org/10.1016/j.ejrad.2011.10.022

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., & Tohka, J. (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI

subjects. *NeuroImage*, *104*, 398–412. http://doi.org/10.1016/j.neuroimage.2014.10.002

Nakatsuka, T., Imabayashi, E., Matsuda, H., Sakakibara, R., Inaoka, T., & Terada, H. (2013). Discrimination of dementia with Lewy bodies from Alzheimer's disease using voxel-based morphometry of white matter by statistical parametric mapping 8 plus diffeomorphic anatomic registration through exponentiated Lie algebra. *Neuroradiology*, *55*(5), 559–566. http://doi.org/10.1007/s00234-013-1138-9

Narasimhamurthy, A. (2005). Theoretical bounds of majority voting performance for a binary classification problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(12), 1988–1995. http://doi.org/10.1109/TPAMI.2005.249

Ortiz, A., Górriz, J. M., Ramírez, J., & Martínez-Murcia, F. J. (2013). LVQ-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease. *Pattern Recognition Letters*, *34*(14), 1725–1733. http://doi.org/10.1016/j.patrec.2013.04.014

Papakostas, G. a., Savio, a., Graña, M., & Kaburlasos, V. G. (2015). A lattice computing approach to Alzheimer's disease computer assisted diagnosis based on MRI data. *Neurocomputing*, *150*, 37–42. http://doi.org/10.1016/j.neucom.2014.02.076

Petrella, J. R., Coleman, R. E., & Doraiswamy, P. M. (2003). Neuroimaging and

Early Diagnosis of Alzheimer Disease : A Look to the Future, 315–336.

Pohjalainen, J., Räsänen, O., & Kadioglu, S. (2015). Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. *Computer Speech & Language*, *29*(1), 145–171. http://doi.org/10.1016/j.csl.2013.11.004

Prati, R. C. (2012). Combining feature ranking algorithms through rank aggregation. *Proceedings of the International Joint Conference on Neural Networks*, 10–15. http://doi.org/10.1109/IJCNN.2012.6252467

Querbes, O., Aubry, F., Pariente, J., Lotterie, J.-A., Démonet, J.-F., Duret, V., … Celsis, P. (2009). Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain : A Journal of Neurology*, *132*(Pt 8), 2036–2047. http://doi.org/10.1093/brain/awp105

Rajapakse, J. C., Giedd, J. N., & Rapoport, J. L. (1997). Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Transactions on Medical Imaging*, *16*(2), 176–186. http://doi.org/10.1109/42.563663

Ram, J., Segovia, F., & Chaves, R. (2009). Multivariate Approaches for Alzheimer ' s Disease Diagnosis Using Bayesian Classifiers, 190–193.

Ramírez, J., Górriz, J. M., Segovia, F., Chaves, R., Salas-Gonzalez, D., López, M., … Padilla, P. (2010). Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image

classification. *Neuroscience Letters*, *472*(2), 99–103. http://doi.org/10.1016/j.neulet.2010.01.056

Ruiz, R., Riquelme, J., & Aguilar-Ruiz, J. (2003). Fast feature ranking algorithm. *Knowledge-Based Intelligent …*, 325–331. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-45224-9_46

Savio, A., García-Sebastián, M. T., Chyzyk, D., Hernandez, C., Graña, M., Sistiaga, A., … Villanúa, J. (2011). Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI. *Computers in Biology and Medicine*, *41*(8), 600–610. http://doi.org/10.1016/j.compbiomed.2011.05.010

Segovia, F., Górriz, J. M., Ramírez, J., Salas-González, D., & Álvarez, I. (2013). Early diagnosis of Alzheimer's disease based on Partial Least Squares and Support Vector Machine. *Expert Systems with Applications*, *40*(2), 677–683. http://doi.org/10.1016/j.eswa.2012.07.071

Seixas, F. L., Zadrozny, B., Laks, J., Conci, A., & Muchaluat Saade, D. C. (2014). A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. *Computers in Biology and Medicine*, *51*, 140–58. http://doi.org/10.1016/j.compbiomed.2014.04.010

Shao, J., Myers, N., Yang, Q., Feng, J., Plant, C., Böhm, C., … Sorg, C. (2012). Prediction of Alzheimer's disease using individual structural connectivity networks. *Neurobiology of Aging*, *33*(12), 2756–2765.

http://doi.org/10.1016/j.neurobiolaging.2012.01.017

Silveira, M., & Marques, J. (2010). Boosting Alzheimer Disease Diagnosis using PET images. http://doi.org/10.1109/ICPR.2010.626

Sima, C., Braga-Neto, U., & Dougherty, E. R. (2005). Superior feature-set ranking for small samples using bolstered error estimation. *Bioinformatics*, *21*(7), 1046–1054. http://doi.org/10.1093/bioinformatics/bti081

Sima, C., Braga-Neto, U. M., & Dougherty, E. R. (2011). High-dimensional bolstered error estimation. *Bioinformatics*, *27*(21), 3056–3064. http://doi.org/10.1093/bioinformatics/btr518

Slavkov, I., Zenko, B., & Dzeroski, S. (2010). Evaluation Method for Feature Rankings and their Aggregations for Biomarker Discovery. *Journal of Machine Learning Research - Proceedings Track*, *8*, 122–135. Retrieved from http://dblp.uni-trier.de/db/journals/jmlr/jmlrp8.html#SlavkovZD10

Son, J. H., Han, D. H., Min, K. J., & Kee, B. S. (2013). Correlation between gray matter volume in the temporal lobe and depressive symptoms in patients with Alzheimer's disease. *Neuroscience Letters*, *548*, 15–20. http://doi.org/10.1016/j.neulet.2013.05.021

Song, X., & Chen, N. K. (2014). A SVM-based quantitative fMRI method for resting-state functional network detection. *Magnetic Resonance Imaging*, *32*(7), 819–831. http://doi.org/10.1016/j.mri.2014.04.004

Stebbins, G., & Murphy, C. (2010). Diffusion tensor imaging in Alzheimer's disease and mild cognitive impairment. *Behav Neurol*, *21*(1), 39–49. http://doi.org/10.3233/BEN-2009-0234.Diffusion

Stoeckel, J., & Fung, G. (2005). SVM Feature Selection for Classification of SPECT Images of Alzheimer's Disease using Spatial Information Computer Aided Diagnosis Computer Aided Diagnosis. *Proceedings of the Fifth IEEE International Conference on Data Mining*.

Stoeckel, J., Migneco, O., Koulibaly, P. M., Robert, P., Ayache, N., & Darcourt, J. (2004). Classification of SPECT Images of Normal Subjects versus Images of Alzheimer ' s Disease Patients, 666–674.

Strobl, C., Boulesteix, A., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures : Illustrations , sources and a solution, *21*. http://doi.org/10.1186/1471-2105-8-25

Svm, C. (2008). Automatic Computer Aided Diagnosis Tool using, 4392–4395.

Takao, H., Hayashi, N., & Ohtomo, K. (2015). Brain morphology is individual-specific information. *Magnetic Resonance Imaging*. http://doi.org/10.1016/j.mri.2015.03.010

Tang, L., Peng, S., Bi, Y., Shan, P., & Hu, X. (2014). A new method combining LDA and PLS for dimension reduction. *PLoS ONE*, *9*(5). http://doi.org/10.1371/journal.pone.0096944

Tang, L., Wen, Y., Zhou, Z., von Deneen, K. M., Huang, D., & Ma, L. (2013). Reduced field-of-view DTI segmentation of cervical spine tissue. *Magnetic Resonance Imaging*, *31*(9), 1507–1514. http://doi.org/10.1016/j.mri.2013.07.003

Teipel, S. J., Born, C., Ewers, M., Bokde, A. L. W., Reiser, M. F., Möller, H.-J., & Hampel, H. (2007). Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *NeuroImage*, *38*(1), 13–24. http://doi.org/10.1016/j.neuroimage.2007.07.008

Tohka, J., Zijdenbos, A., & Evans, A. (2004). Fast and robust parameter estimation for statistical partial volume models in brain MRI. *NeuroImage*, *23*(1), 84–97. http://doi.org/10.1016/j.neuroimage.2004.05.007

Vemuri, P., Gunter, J. L., Senjem, M. L., Whitwell, J. L., Kantarci, K., Knopman, D. S., … Jack, C. R. (2008). Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. *NeuroImage*, *39*(3), 1186–1197. http://doi.org/10.1016/j.neuroimage.2007.09.073

Vemuri, P., & Jack, C. R. (2010). Role of structural MRI in Alzheimer's disease. *Alzheimer's Research & Therapy*, *2*(4), 2:23. http://doi.org/10.1186/alzrt47

Wang, D., Zhang, H., Liu, R., Lv, W., & Wang, D. (2014). T-Test feature selection approach based on term frequency for text categorization. *Pattern Recognition Letters*, *45*(1), 1–10. http://doi.org/10.1016/j.patrec.2014.02.013

Wang, W., Wang, X., Feng, D., Liu, J., Han, Z., & Zhang, X. (2014). Exploring Permission-Induced Risk in Android Applications for Malicious Application Detection, *9*(11), 1869–1882.

Watanabe, H., Ono, M., & Saji, H. (2015). Novel PET / SPECT Probes for Imaging of Tau in Alzheimer ' s Disease, *2015*.

Wee, C. Y., Yap, P. T., Zhang, D., Denny, K., Browndyke, J. N., Potter, G. G., … Shen, D. (2012). Identification of MCI individuals using structural and functional connectivity networks. *NeuroImage*, *59*(3), 2045–2056. http://doi.org/10.1016/j.neuroimage.2011.10.015

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. a., Cairns, N. J., Green, R. C., … Trojanowski, J. Q. (2015). 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's and Dementia*, *11*(6), e1–e120. http://doi.org/10.1016/j.jalz.2014.11.001

Westman, E., Cavallin, L., Muehlboeck, J. S., Zhang, Y., Mecocci, P., Vellas, B., … Wahlund, L. O. (2011). Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease. *PLoS ONE*, *6*(7). http://doi.org/10.1371/journal.pone.0022506

Westman, E., Muehlboeck, J. S., & Simmons, A. (2012). Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *NeuroImage*, *62*(1), 229–238.

http://doi.org/10.1016/j.neuroimage.2012.04.056

Xia, Y., Ieee, M., Wen, L., Ieee, M., Eberl, S., Ieee, M., … Preparation, A. D. (2008). Genetic Algorithm-Based PCA Eigenvector Selection and Weighting for automated identification of Dementia using FDG-PET Imaging, 4812–4815.

Xu, L., Wu, X., Chen, K., & Yao, L. (2015). Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment. *Computer Methods and Programs in Biomedicine*, (19), 1–9. http://doi.org/10.1016/j.cmpb.2015.08.004

Xue, T., Bai, L., Chen, S., Zhong, C., Feng, Y., Wang, H., … Liu, Y. (2011). Neural specificity of acupuncture stimulation from support vector machine classification analysis. *Magnetic Resonance Imaging*, *29*(7), 943–950. http://doi.org/10.1016/j.mri.2011.03.003

Yan, W. (2007). Fusion in multi-criterion feature ranking. *Information Fusion, 2007 10th International Conference on*, 1–6. http://doi.org/10.1109/ICIF.2007.4408064

Yan, Z., Wang, Z., & Xie, H. (2008). The application of mutual information-based feature selection and fuzzy LS-SVM-based classifier in motion classification. *Computer Methods and Programs in Biomedicine*, *90*(3), 275–284. http://doi.org/10.1016/j.cmpb.2008.01.003

Yanxi Liu, C. C. M. (2004). Discriminative MR Image Feature Analysis for

Automatic Schizophrenia and Alzheimer ' s Disease Classification. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*, 393–401.

Yu, S.-N., & Lee, M.-Y. (2012). Conditional mutual information-based feature selection for congestive heart failure recognition using heart rate variability. *Computer Methods and Programs in Biomedicine*, *108*(1), 299–309. http://doi.org/10.1016/j.cmpb.2011.12.015

Zhang, D., Wang, Y., Zhou, L., Yuan, H., & Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*, *55*(3), 856–867. http://doi.org/10.1016/j.neuroimage.2011.01.008

Zhang, J. Z. J., Yan, B. Y. Bin, Huang, X. H. X., Yang, P. Y. P., & Huang, C. H. C. (2008). The Diagnosis of Alzheimer's Disease Based on Voxel-Based Morphometry and Support Vector Machine. *2008 Fourth International Conference on Natural Computation*, *2*, 197–201. http://doi.org/10.1109/ICNC.2008.804

Zhang, L., Song, M., Liu, X., Bu, J., & Chen, C. (2013). Fast multi-view segment graph kernel for object classification. *Signal Processing*, *93*(6), 1597–1607. http://doi.org/10.1016/j.sigpro.2012.05.012

Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A., & Liu, H. (2010). Advancing Feature Selection Research. *ASU Feature Selection Repository Arizona State University*, 1 – 28. Retrieved from

http://featureselection.asu.edu/featureselection_techreport.pdf

Zhou, N., & Wang, L. (2007). A Modified T-test Feature Selection Method and Its Application on the HapMap Genotype Data. *Genomics, Proteomics and Bioinformatics*, *5*(3-4), 242–249. http://doi.org/10.1016/S1672-0229(08)60011-X

Zhou, Q., Goryawala, M., Cabrerizo, M., Wang, J., Barker, W., Loewenstein, D. a, … Adjouadi, M. (2014). An Optimal Decisional Space for the Classification of Alzheimer's Disease and Mild Cognitive Impairment. *Biomedical Engineering, IEEE Transactions on*, *61*(8), 2245–2253. http://doi.org/10.1109/TBME.2014.2310709