

# **Detection of Alzheimer's Disease using 3D MRI Based on Key Slices Selected**

**Masoud Moradi**

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

Master of Science  
in  
Electrical and Electronic Engineering

Eastern Mediterranean University  
September 2017  
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

---

Assoc. Prof. Dr. Ali Hakan Ulusoy  
Acting Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science 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 for the degree of Master of Science in Electrical and Electronic Engineering.

---

Prof. Dr. Hasan Demirel  
Supervisor

---

Examining Committee

1. Prof. Dr. Hasan Demirel

---

2. Assoc. Prof. Dr. Önsen Toygar

---

3. Asst. Prof. Dr. Rasime Uygurođlu

---

## ABSTRACT

Alzheimer's disease (AD) is one of the most common irreversible dementia disease affecting mostly old people specially in older adulthood. This disease corresponds to a particular group of aging dementia which deteriorates long-term and short-term memory, behavior and thinking. Currently there is no treatment for AD, but early detection of AD can help finding out mechanisms of AD and make better life for patients who suffer from AD.

In this study we detect Alzheimer's disease (AD) subjects among elderly cohorts including health and AD subjects. One of the main issues of automatic AD classification is feature extraction in high-dimensional feature space. This thesis proposes new feature extraction methods for high-dimensional pattern recognition problem aimed at accurate detection of AD. The proposed methods use information from three dimensional magnetic resonance imaging (MRI) brain data with 2D slices in three orthogonal directions. The proposed method includes the calculation of Fisher Criterion between the AD and HC groups in order to select key-slices in the coronal, sagittal and axial directions. The preprocessing phase involve region detection to segment region of interest (ROI) based on displacement field (DF) method. Then we utilized energy, contrast and homogeneity metrics along with feature vectors generated by PCA and probability distribution function (PDF) methods in feature extraction phase for each key-slice selected in the earlier phase. Features coming from each key-slice are combined through feature fusion for improved accuracy. Experimental results show fusion method that used with brain mask give us the higher or comparable results compared with other feature extraction techniques in the literature.

**Keywords:** Alzheimer's disease, MRI, region of interest, Statistical feature extraction, data fusion, classification, support vector machine

## ÖZ

Alzheimer hastalığı (AD) çoğunlukla yaşlıları özellikle yaşlılık döneminde etkileyen, geri dönüşü olmayan bunama hastalıklarından biridir. Bu hastalık, uzun süreli ve kısa süreli hafızayı, davranış ve düşünceyi bozan belirli yaşlanma bunama grubuna karşılık gelir. Şu anda AD için herhangi bir tedavi yoktur, ancak AD'in erken teşhisi, AD'in mekanizmalarını bularak AD'den mustarip hastalar için daha iyi yaşam sağlamak için yardımcı olabilir.

Bu çalışmada, yaşlı kitlelerden alınan sağlıklı ve hasta (AD) görsel verileri kullanılarak Alzheimer hastalığı (AD) olanları tespit edilebilmektedir. Otomatik AD sınıflandırmasının ana konularından biri de, yüksek boyutlu öznitelik uzayında öznitelik çıkarımıdır. Bu tez ile AD'nin doğru bir şekilde sezilmesini amaçlayan yüksek boyutlu tanıma problemi için yeni öznitelik çıkarım yöntemleri önerilmektedir. Önerilen yöntemler üç orthogonal doğrultudaki üç boyutlu manyetik rezonans görüntüleme (MRI) beyin verilerinden gelen bilgilerden çıkarılan anahtar kesitleri kullanmaktadır. Önerilen yöntem, koronal, sagittal ve aksenel yöndeki anahtar kesitleri seçmek için AD ve HC grupları arasındaki Fisher Kriteri hesaplanmasını içermektedir. Önişleme aşaması, kesitler arası yer değiştirme alanı (DF) yöntemine dayalı olarak ilgi bölgesi (ROI) için bölge algılamasını içerir. Algılanan bu bölgeden beyin maskesi çıkarımı yapılır. Sonrasında, enerji, kontrast ve homojenlik ölçümleri yanında, PCA tarafından üretilen öznitelik vektörleri ile olasılık dağılım fonksiyonu (PDF) yöntemlerinden önceki aşamada seçilen her bir anahtar kesit için çıkarılan öznitelikleri kullandık. Her bir anahtar kesitinden gelen öznitelikler, performansın artırılması için öznitelik düzeyi kaynaşımı ile birleştirilmiştir. Deneysel sonuçlar, literatürdeki diğer

öznitelik çıkarım teknikleriyle karşılaştırıldığında beyin maskesi kullanılan kaynaşım yönteminin daha yüksek veya karşılaştırılabilir sonuçlar verdiğini göstermektedir.

**Anahtar Kelimeler:** Alzheimer hastalığı, MRI, ilgi alanı, istatistiksel öznitelik çıkarma, veri kaynaşımı, sınıflandırma, destek vektör makinesi

## **ACKNOWLEDGMENT**

I would like to express my appreciations to my supervisor Prof. Dr. Hasan Demirel for giving me this opportunity to work on this thesis. His guidance, encouragements, and patience will always be appreciated. It was a pleasure and honor for me to work with him.

# TABLE OF CONTENTS

ABSTRACT.....	iii
ÖZ.....	v
ACKNOWLEDGMENT.....	vii
LIST OF TABLES.....	xi
LIST OF FIGUERS.....	xii
LIST OF SYMBOLS AND ABBREVIATIONS.....	xiii
1 INTRODUCTION.....	1
1.1 Introduction.....	1
1.2 Problem Definition.....	2
1.3 Thesis Objectives.....	3
1.4 Thesis Contributions.....	3
1.4 Thesis Overview.....	4
2 LITERATURE REVIEW.....	5
2.1 Introduction.....	5
2.2 Alzheimer Disease.....	5
2.3 Magnetic Resonance Imaging (MRI).....	7
2.4 Detection of AD Based on 2D MRI Scans.....	9
2.5 Detection of AD Based on 3D MRI Scans.....	9
3 METHODOLOGY.....	11
3.1 Brain MRI Database.....	11
3.2 Data Preprocessing.....	12
3.3 Accuracy, Sensitivity and Specificity.....	12
3.4 Receiver Operating Characteristic (ROC) Curve.....	15



3.6 Principle Component Analysis (PCA).....	17
3.7 Cross Validation .....	19
4 CLASSIFICATION TECHNIQUES.....	21
4.1 Introduction.....	21
4.2 Support Vector Machine Classifier .....	22
4.3 K-Nearest Neighbor Classifier .....	23
4.4 Decision Trees Classifier.....	24
5 DETECTION OF ALZHEIMER DISEASE BASED ON KEY SLICES AND DATA FUSION.....	26
5.1 Introduction.....	26
5.2 Key-Slice Selection .....	28
5.3 Region Detection .....	28
5.4 Feature Extraction.....	30
5.4.1 Feature Extraction Based on PDF.....	30
5.4.2 Feature Extraction Based on Statistical Parameters .....	31
5.4.3 Feature Extraction Based on PCA .....	32
5.5 Data Fusion.....	33
5.6 Classification .....	34
6 RESULTS AND DISCUSSIONS .....	35
6.1 Key Slice Selection.....	35
6.2 Region Detection .....	40
6.3 Feature Extraction and Classification Comparison .....	46
7 CONCLUSION AND FUTURE WORK .....	49
7.1 Conclusion .....	49
7.2 Future Work.....	50

REFERENCES ..... 51

## LIST OF TABLES

Table 3.1: Subject demographics and dementia status.....	11
Table 3.2:Relationship Between the Area Under the ROC Curve and Diagnostic Accuracy .....	17
Table 6.1: The Accuracy of Our Propose Method Without Using Mask.....	47
Table 6.2: The Accuracy of Our Propose Method With Mask .....	47
Table 6.3: Comparison Our Method with Other Methods.....	48

## LIST OF FIGUERS

Figure 2.1: Illustration of brain atrophy using sMRI in 3 major axis .....	8
Figure 3.1: a) Raw Image, b) Image Normalization, c) Brain Mask .....	12
Figure 3.2: Dependency of True Positive, True Negative, False Negative and False Positive.....	14
Figure 3.3: Receiver Operating Characteristic Curve [33]. .....	16
Figure 3.4: 10 fold cross-validation .....	20
Figure 4.1: Support Vector Machine classifier .....	22
Figure 4.2: K-Nearest Neighbor Classifier .....	23
Figure 5.1: Block diagram of proposed AD detection method. ....	27
Figure 5.2: The diagram of displacement field generation. ....	29
Figure 6.1: Curve of Fisher Criteria for coronal direction .....	36
Figure 6.2: Curve of Fisher Criteria for sagittal direction .....	37
Figure 6.3: Curve of fisher criteria for axial direction .....	37
Figure 6.4: Key-Slice of coronal from 60 to 150 with up-sampling 10.....	38
Figure 6.5: Key-Slice of sagittal from 40 to 130 with up-sampling 10 .....	39
Figure 6.6: Key-Slice of axial from 60 to 100 with up-sampling 10 .....	40
Figure 6.7: Flowchart of the proposed Region Detection Method.....	42
Figure 6.8: Brain Mask of Axial View from 60 to 100 with Up-Sampling 10.....	43
Figure 6.9: Brain Mask of Coronal View From 60 to 150 with Up-Sampling 10.....	44
Figure 6.10: Brain Mask of Sagittal View From 60 to 150 with Up-Sampling 10.....	45

## LIST OF SYMBOLS AND ABBREVIATIONS

3-D	Three Dimension
2-D	Two Dimension
$\mu$	Mean
$C_i$	Class number
K	Fold
$S_i$	Support Vector
$S_B$	between-class scatter matrix
$S_W$	within-class scatter matrix
$\bar{x}$	Mean
ACC	Accuracy
AD	Alzheimer's Disease
ANN	Artificial Neural Network
AUC	Area Under a Curve
CAD	Computer-Aided Diagnosis
CDR	Clinical Dementia Ratio
CFS	Cerebrospinal Fluid
CT	X-ray computed Tomography

CV	Cross Validation
DF	Displacement Field
DT	Decision Tree
DWT	Discrete Wavelet Transform
ECH	Energy, Contrast and Homogeneity
FC	Fisher Criterion
FP	False Positive
FPANN	Forward Back-propagation Artificial Neural Net-work
FPR	False-Positive Ratio
FN	True Negatives
Fmri	Functional Magnetic Resonance Imaging
GR	Gradient Recalled
GM	Gray Matter
H	Histogram
HC	Healthy Control
IG	Information Gain
KNN	K-Nearest Neighbor
KS	Key Slice
MMSE	Mini Mental State Examination

MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
OASIS	Open Access Series of Imaging Studies
PC	Principal Component
PCA	Principal Component Analysis
PDF	Probability Distribution Function
PET	Positron Emission Tomography
ROC	Receiver Operating Characteristic
ROI	Regions of Interest
SD	Statistical Dependency
SE	Standard Error
SEN	Sensitivity
SNR	Signal to Noise Ratio
SPE	Specificity
SPECT	Single Photon Emission Computed Tomography
sMRI	Structural Magnetic Resonance Imaging
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

WB	Whole Brain
WM	White Matter



# Chapter 1

## INTRODUCTION

### 1.1 Introduction

Alzheimer's Disease (AD) is an illness affecting mostly older people particularly those in their 60's [1]. It's a progressive dementia which causes changes in behavior, loss of memory, thinking and language skills [2]. The AD syndrome gradually worsens over time and eventually meddles with a patient's daily normal life and eventually kills the affected victim. No cure has been discovered yet for this illness. Symptoms of AD has interested researchers from everywhere throughout the world due to its significance and impact on society [3]. All through the cycle of AD, side effects of AD may turn out to be more serious. From 2006 the overall global population that suffered from AD is estimated to be 26.6 million [4] [5]. It is estimated by 2050 the population of people who will be affected and stricken by AD to increase to 106 million and around 40 percent of these cases require intensive care [6].

Since the world is constantly developing, AD has brought about a greater negative impact on different societies of the world at large [7]. In China, AD represents the greater part of SD, which removes an aggregate financial drop of more than 80 billion yuan consistently, and therefore it's in charge of almost 60 billion yuan in human service expenses each year [8]. In the United States, social insurance covering people

suffering with AD generally amounts to almost \$100 billion yearly and is anticipated to cost a trillion every year by twenty-fifty [9].

These days, it is valuable to improve new and reliable procedures for diagnosing AD, which is likewise fundamental for the cure and administration of expediting the disintegration of Alzheimer Disease [10]. A 3D scan of the entire cerebrum gets to be distinctly adequate and moderate with late advances in neuroimaging innovation, particularly through the assistance of the most renowned imaging procedure: Magnetic Resonance Imaging (MRI). With its enormous pixel focus magnetic resonance (MRI) pictures, the precision of detecting Alzheimer Disease is enormously improved. For distinguishing AD, MR images now assumes the basic part from the normal elder controls (NC) [11].

A recent research study has discovered different processes of distinguishing stages of Alzheimer disease. Majority of these researched processes use three steps: (1) The first step is feature extraction which extract best features that detect deformed brains from healthy brains. (2) Feature selection, to select best features and reasonable numbers of it leads to reduce the size of features. (3) Classification, to make a classification using extracted features for distinguishing Alzheimer Disease.

## **1.2 Problem Definition**

High dimensional classification techniques to achieve high accuracy are imperative for achievement of numerous applications, particularly in automated detection of patient suffering from Alzheimer Disease. A number of neuroimaging studies have been introduced by several high-dimensional pattern recognition procedures [12] [13] [14]. A feature-extraction procedure from a greater-dimensional data classification is

improvised to decrease the computational cost and also improve the accuracy. This procedure is very successful in processing the conclusive results. In this thesis, three most imperative and effective feature extraction methods have been presented to solve the problem of high-dimensional pattern recognition in Alzheimer Disease.

### **1.3 Thesis Objectives**

The main potential attainment of machine learning for automated detection of patients who endure AD is high dimensional classification technique. Feature selection and dimension reduction technique from high dimensional data is one of the most important topics in pattern recognition and data mining. Over the last few years, many studies successfully verified high-dimensional pattern recognition in a number of neuroimaging studies. In this thesis, we introduce a new way for feature extraction in high-dimensional detection of AD.

### **1.4 Thesis Contributions**

A major target of machine learning derived from high-dimensional detection procedures is a major automated classification of patients suffering from Alzheimer's disease (AD). The feature-extraction method from high-dimensional data is the biggest issue of the automated classification. In a number of neuroimaging studies, a number of researching studies investigated high-dimensional pattern detection approaches. In this thesis, we present imperative feature extraction procedures in high-dimensional recognition of Alzheimer Disease. Main contributions of this study can be listed as:

- 1- Using fisher criteria for selecting key slices (KS)
- 2- Generating 2D mask for segmentation of region of interest (ROI)

- 3- Introducing a new method for feature extraction based on data fusion to develop classification enactment by merging PDF features, Eigen brain values and statistical features.

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

## **1.4 Thesis Overview**

The recent studies in AD detection of chapter 2 give descriptions of literature reviews. The methodology used in chapter 3 presents our thesis including image data pre-processing, definition of accuracy, principal component analysis, sensitivity, specificity, receiver operating characteristic and cross-validation. In chapter 4 we introduce some classification methods. Chapter 5 is describing the state-of-the-art methodology key-slice selection from 3D MR images and also the region of detection and then describe some of the methods of feature extraction. Additionally, this chapter presents the data fusion techniques presented to improve the detection performance. Chapter 6 presents results from chapter 5, chapter 4 and chapter 3 including key-slice selection, region detection and classification performance and also compares the conclusion with other research studies and related articles. In chapter 7 the thesis conclusion is presented according to the discussions and interpretations contributed in this study. Furthermore, this research incorporates recommendation for the enhancement of this study.

## Chapter 2

### LITERATURE REVIEW

#### 2.1 Introduction

Alzheimer's disease (AD) is a brain disorder that affects the aging process and has a negative influence on some parts of the brain that are responsible for memory, learning, and higher executive functioning [4]. In the history of medical research for the cure of AD, there hasn't been any discovery until recently [7]. Early and correct diagnosis of AD is helpful for the development and management of this disease. MRI is a medically renowned procedure that uses a light ray to give high density images of the anatomical structures of the human body and it's also used to detect and reveal AD during the early diagnoses [5] [10] [15]. In the earlier studies for detecting AD, it was discovered that the capture of the disease works on the 2D or 3D MRI images based on whole images or region of interest (ROI) [12] [16] [17].

#### 2.2 Alzheimer Disease

The first case for Alzheimer's disease was presented by Alois Alzheimer who was a German psychiatrist. He diagnosed one of his patients, a fifty years old woman, and she sadly passed away in 1906, following vague information that came to light regarding the cure of the disease. He published his first paper about Alzheimer's disease in 1906 [18].

Alzheimer's Disease is a common cause of brain damages and mind illnesses and amongst many it's attributes and symptoms include signs manifested by difficulties in

long and short memory, disturbances in language cognition, psychological and psychiatric changes and also disturbance of normal daily activities. It is estimated that AD affects around 6% of the entire population aged more than 60 years old and also devastatingly disables the affected patient the older they get. Patients with AD are often identified and are then instantly taken into the intensive care unit, whereby diagnostic and management challenges of patients are presented. The benefits of early investigation and diagnosis of AD includes the initiation of identified symptoms treatment and initiation of an ongoing prolonged support, with the additional remedy of affected facets and areas of life of the diagnosed patients [19].

Currently AD is diagnosed through clinical tests and evaluation of realization, and impairment behavior of a suffering patient is deciphered as an indicator appearing in the last stages of the disease. Neuroimaging is regarded as the best early detection procedure of AD because it indicates the restructuring and functional activities in the brain. In this era, Neuroimaging, data analysis and tendency of using local features have become controversial medical procedures. The Magnetic Resonance Imaging (MRI), X-Ray computed tomography (CT), Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), are procedures usually used medically for early detection of AD [20] [21] [22]. According to doctors and medical scientists, the best procedure for detecting AD when having to compare between Neuroimaging and MR imaging is the highly preferred because MRI images have high spatial resolution imaging of organs and soft tissue, good contrast and high accessibility [23] [24].

## 2.3 Magnetic Resonance Imaging (MRI)

The occurrence of nuclear magnetic resonance (NMR), the microscopic nitty-gritty outline of MR was purposed in 1945 [25] [26]. The merging of the radiofrequency engineering and the accessibility of the new recurrence sources contributed to the birth and showcasing of the Nuclear Magnetic Resonance(NMR). The information about NMR imaging was initially publicly disclosed in 1973, before the end of a decade the first human being had already started examining the first images of which a go ahead for the first production of the images was enacted. Contrasted with MR pictures from different regions of the body, images of the head gave outstanding anatomical feature and robust gray/white matter differentiate. Stream delicate systems created in the mid-1980s measured blood stream velocity; these were the principal MR images assumed not to be simply basic in nature. The fantastic presentation of functional MR imaging (fMRI) a decade ago developed MRI to such an extent that in present MRI, the explanation of structural MRI looks to have moved meaning of the not utilitarian MRI [27] [28].

For differentiating between functional and structural imaging is hard since structures and functions are most of the times, inseparable in the brain. Meanings of functional imaging are broad and often vary, and therefore will always dependably be subjective to a specific degree. On the fundamental of organic contemplations, functional imaging can be viewed as the technique giving dynamic physiological data, while structural imaging gives static anatomical data. fMRI accordingly incorporates BOLD (blood oxygen level ward) system. fMRI can be taken to be stage differentiate stream estimations. Magnetic resonance angiography is difficult to describe and give meaning, but since its purpose is to define how well the vessels carry blood, it can be

taken to be a functional technique. Spectroscopy and chemical shift imaging are set out to calculate chemical concentrations, and these should differ from other MR techniques. Spectroscopy mostly is structural/static in physique, although some spectroscopy operations studies have been conducted [27].

It's most definitely further away from the scope of any research to measure different methods and functions of MRI structure. shows brain atrophy using sMRI from various point of view.

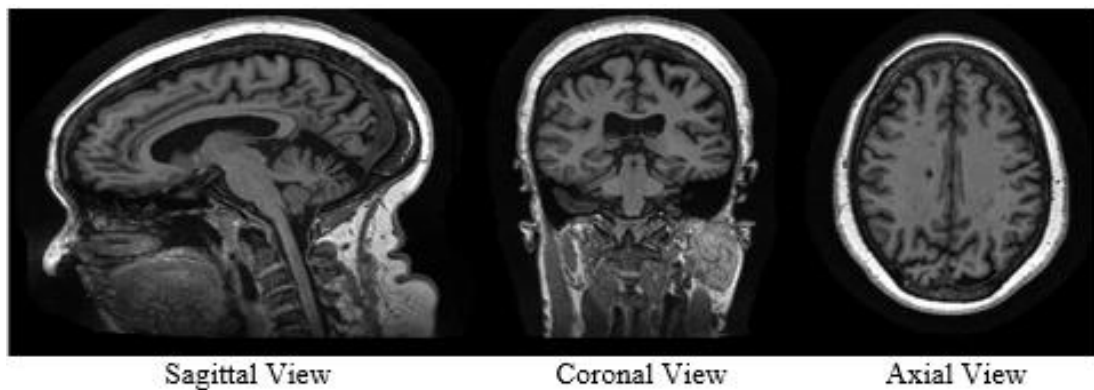


Figure 2.1: Illustration of brain atrophy using sMRI in 3 major axis

Spatial resolution and signal to noise ratio (SNR) are the two main properties that manage image quality. The number of image elements (pixels) in the frequency and phase encoding directions and Through-plane Resolution by the slice thickness determine In-plane Resolution. SNR is determined by pixel size, slice thickness, scan time and the order utilized. Image quality is determined by total scan time which is chiefly influenced by patient's ability to keep still. Detection of AD is based on two methodologies: two dimensional MRI scans demand and three dimensional MRI scans [27].



## **2.4 Detection of AD Based on 2D MRI Scans**

Some of the studies such as Yudong Zhang & Shuihua Wang they work on 2 dimensional MRI scans and find ROI of the brain and then use SVM classifier [29]. El-Dahshan, Hosny and Salem suggested the utilization of two-dimensional DWT (2D-DWT) with principal component analysis (PCA) in the first stage, and use forward back-propagation artificial neural network (FPANN) and k-nearest neighbors (KNN) in the next stage [30]. S. Chaplot and L.M. Patnaik used coefficients of discrete wavelet transform (DWT) as input to neural network self-organizing maps and support vector machine to classify MR images of the human brain [31]. Zhang and W. Jagannathan utilized coefficients of discrete wavelet transform (DWT) and they placed the features in the kernel support vector machine (KSVM) classifier.

## **2.5 Detection of AD Based on 3D MRI Scans**

Analysis “whole-brain (WB)” is a new kind of method, that has gained familiarity because it includes all voxels in the WB. It doesn't require to divide the brain, and for classification task there is not necessary a brain mask. The curse of dimensionality is the major disadvantage, which can be mitigated via the usage of a computer which is cheaply accessible nowadays. The overall brain analysis intimately relies on computational algorithms, and therefore only computer scientists can accomplish it, only after a go ahead from the physicians to determine the data to be AD or NC. The WB analysis provides the WB as a ROI generally.

Some of the studies such as Zhang Y, Wang S, Phillips P, they proposed 3D discrete wavelet transform (3D-DWT) for the study of the WB to extract wavelet coefficients and the volumetric image [16]. Plant et al applied brain region cluster (BRC), and

proposed the utilize of information gain (IG) to calculate the ROI of a voxel. They used Bayes statistics, SVM, and voting feature intervals (VFI) to classify patterns [32].

## Chapter 3

### METHODOLOGY

#### 3.1 Brain MRI Database

The public dataset Open Access Series of Imaging studies (OASIS, downloaded from <http://oasis-brains.org/> ) is collected from, of which incorporates 416 subjects with dataset age from 18 to 96 respectively. All persons are right-handed. Because some of the samples are disintegrated with some of the data, we've chosen the samples that had all data information. At the end, we've chosen 127 samples (30 ADs and 97 HCs) from the dataset. Table 4.1 below consist of a demographic report. The Clinical Dementia Rating database was explained as the target (label) after the common convention. Subjects with missing records or the ones which were less than 60 years old were removed.

Table 4.1: Demographic rank of subjects.

Characteristic	AD	HC
Subject	30	97
Age	77.75 $\pm$ 6.98	75.91 $\pm$ 8.96
Sex	10/20	27/72
Academic status	2.57 $\pm$ 1.33	3.26 $\pm$ 1.33
Socioeconomic status	2.87 $\pm$ 1.28	2.51 $\pm$ 1.11
MMSE	21.67 $\pm$ 3.78	28.95 $\pm$ 1.22

### 3.2 Data Preprocessing

The 3D MR brain images were motion-adjusted and spatially co-registered to the Talairach space in order to increase the SNR and generate an averaged image, and then thereafter were covered with brain mask. An example of preprocessing on 3D images with resolution of  $1\text{mm}\times 1\text{mm}\times 1.25\text{mm}$  is shown below on Figure 4-1. The 3D images of three scans were registered via the motion-adjusted, and then used to create an image in the original procurement space with a resize to  $1\text{mm}\times 1\text{mm}\times 1\text{mm}$ . The averaged image was then standardized to the Talairach coordinate space, and after that the brain was separated from the skull.

### 3.3 Accuracy, Sensitivity and Specificity

A faultless indicative treatment has the ability to purely diminish the subjects with and without illness. In figure 3.1 the standards of a rounded experiment which are above the line always show the affected areas, whereas the values under the line show the unaffected areas. It's unlikely in real life to witness such biased tests because detection procedures make miniscule difference between co-related subjects with and without illness.

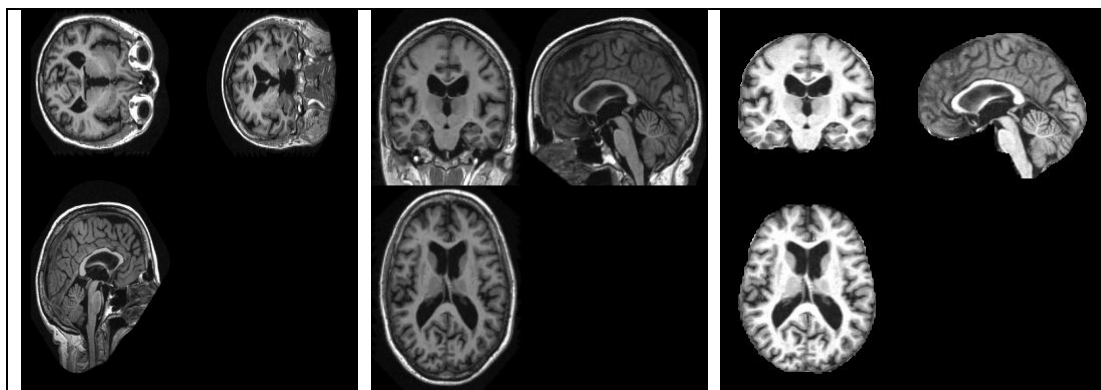


Figure 3.1: a) Raw Image, b) Image Normalization, c) Brain Mask

Standards above the division line aren't always exhibiting the disease because subjects without the disease can also have greater values in certain times. Such major standards of particular criterion of interest are referred to as false positive values (FP). Whilst other standards below the division line are usually located in subjects without illness. However, subjects affected by the illness can have those too. Those standards are false negative values (FN). The cross checked subjects affected or unaffected by the illness are separated by a division line in 4 subgroups considering these significant values of interest.

True Positive (TP): The number of Alzheimer Disease can correctly distinguish as AD.

False Positive (FP): The number of HC can correctly distinguish as HC.

True Negative (TN): The number of subjects with AD wrongly distinguish as HCs.

False Negative (FN): The number of HCs wrongly distinguish as AD.

A detailed Venn diagram of the arrangement of the true negative, false negative, true positive and false positive outcomes for a test to identify the disease (Figure: 3.2). The area where found and actual search results (TP) includes the correct search results. The TN indicate area that incorrectly detected the search result. The FP area refers to search results that were found, but do not include an actual search result. The FN area refers to research that are actual search results, but were not found.

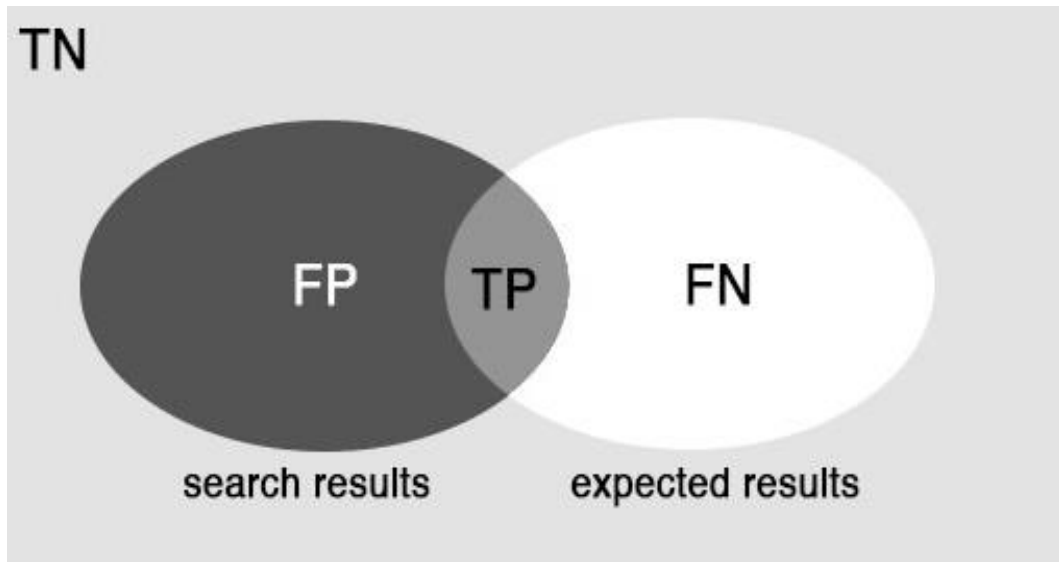


Figure 3.2: Dependency of True Positive, True Negative, False Negative and False Positive

The quantification of true positives that are aligned appropriately with a detection test are defined through sensitivity in percentage

$$\text{sensitivity or true positive rate (TPR)} = \frac{TP}{TP+FN} \quad (3.1)$$

Sensitivity is defined as the likelihood of obtaining a positive test result in a subject with the illness.

Therefore, it exhibits the likelihood of an assessment to detect an illness.

A degree of an indicative test accuracy associated to sensitivity is Specificity. It is defined as a blend of subjects with an absence of illnesses and an overall negative outcome in total subjects affected by the illness.

$$\text{specificity (SPC) or true negative rate} = \frac{TN}{TN+FP} \quad (3.2)$$

In other words, specificity indicate the likelihood of a negative test outcome in a case of the absence in a disease.

Therefore, we can cleverly guess that specificity conforms to the side of precision that describes the examination ability to decipher subjects without the illness.

Accuracy is expression ratio of correctly classified subjects or true results (TP+TN) in all population or all subjects (TP+TN+FP+FN). It evaluates the degree of veracity of a diagnostic test on a condition [33].

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (3.3)$$

### **3.4 Receiver Operating Characteristic (ROC) Curve**

The sensitivity against specificity standards for every individual division line can be defined as a duo of detection tests. A ROC graph can be represented by exhibiting all existing combinations of duo standards graphically by placing specificity on the x-axis and sensitivity on the y-axis (Figure 3.3). Both TPR and FPF have an equilibrium point of (0, 1) in a perfect circumstance and that could be defined as an idle point or we could define it as a point showing up in the left corner of the ROC space. Since this ideal point exhibits the symptomatic experiment and has both a sensitivity of 100% and specificity of 100% we can then now regard it as an ideal classification [34].

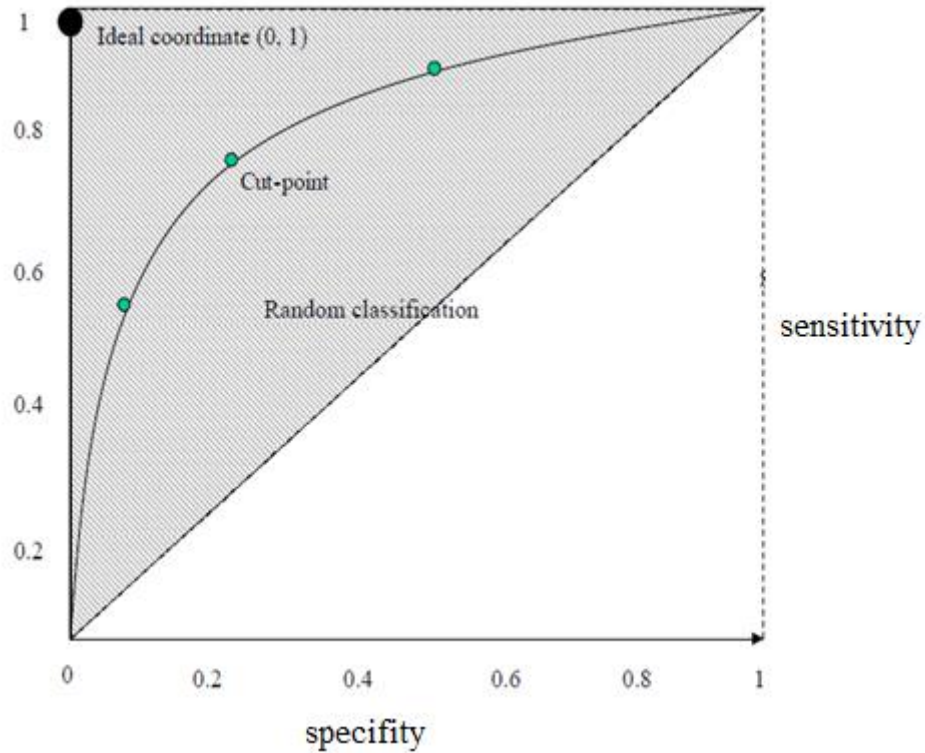


Figure 3.3: Receiver Operating Characteristic Curve [33].

The area under the ROC curve (AUC) as well as its shape assist us in measuring the precision of a diagnostic test. The greater the area, the higher the accuracy of the diagnostic test. AUC of ROC curve can be calculated using this equation (3.4):

$$AUC = \int_0^1 ROC(x)dx \quad (3.4)$$

Whereby  $x = (1 - \text{specificity})$  and  $ROC(x)$  equals sensitivity. The area under the curve serves as a good indicator of the credibility of the test and therefore could have a value of between 0 and 1. A perfect diagnostic has a test value of AUC 1.0 whereas a non-distinguishing test has an area of 0.5. We can now conclude that Table 3.2 below summarizes the relationship between AUC and diagnostic accuracy as follows:



Table 3.2: A summary relationship between AUC and diagnostic accuracy

<b>AUC Range</b>	<b>diagnostic accuracy</b>
0.9-1.0	Outstanding
0.8-0.9	Very well
0.7-0.8	well
0.6-0.7	Satisfying
0.5-0.6	Poor
< 0.5	Invalid

Diagnostic accuracy is generally measured by AUC. It entails less details regarding parameters such as sensitivity and specificity. Two of the tests with the same AUC could give two different output, one with higher specificity and the other with higher sensitivity. Consequently, the results obtained from AUC do not show any predictive values and also do not show any sign of a conclusion of a diagnosis. Diagnostic tests could be generally measured by comparison and also by global measures. By comparing areas below the two ROC curves, we could evaluate which test will give relevant information distinguishing two different scenarios of interest. The comparison should never be based on visible or emotional evaluation (4). To intimately align with the purpose, we use assessment tests which determine the assessment relevance of evaluated difference between two AUC, using a relevant dimension of assessment connotation (P) [33].

### **3.6 Principle Component Analysis (PCA)**

One effective component to reduce the element on a dataset is PCA and it contains a greater amount of variables depending on each other to collectively change into a miniscule amount of new variables. It's a process that converts a dataset to an innovative set of collectively aligned variables in accordance to their variances. The effects of this methodology are orthogonalizing the tools of the input vectors so that they're coherent to one another, ordering orthogonal tools in order for the largest variation to be the first and terminate tools contributing to the last variation in the data

and also normalize input vectors to attain insignificant mean and unity variance before proceeding to perform PCA.

Principle components consist of linear combinations that delegates high-dimensional and innovative 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:

$$X_{n \times m} = \begin{pmatrix} x_{1,1} & \dots & x_{1,m} \\ \vdots & & \vdots \\ x_{n,1} & \dots & x_{n,m} \end{pmatrix} \quad (3.5)$$

Principle components are eigenvectors of the covariance matrix of data  $X$ . The covariance matrix is defined as follow:

$$C_{n \times n} = \begin{pmatrix} c_{1,1} & \dots & c_{1,n} \\ \vdots & \ddots & \vdots \\ c_{n,1} & \dots & c_{n,n} \end{pmatrix} \quad (3.6)$$

where  $c_{j,k}$  is figure out by the following:

$$c_{j,k} = \sum_{i=1}^m \frac{(x_{i,j} - \bar{x}_j)(x_{i,k} - \bar{x}_k)}{m-1} \quad (3.7)$$

where  $\bar{x}_j$  and  $\bar{x}_k$  are the mean of columns  $j$  and  $k$ .  $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$  are ordered eigen-values of covariance matrix. The eigen-vector (i.e.,  $q$ ) of covariance matrix is defined as follow:

$$Cq = \lambda q \quad (3.8)$$

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

$$Q = [q_1 \ q_2 \ \dots \ q_k] \quad (3.9)$$

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

### 3.7 Cross Validation

Cross validation is a technique and sometimes it's called rotational estimation. The whole dataset we called D, was randomly divided to number of fold (divided to K), then we have K subset ( $D_1, D_2, \dots, D_K$ ) with approximately same size.

For the K-fold cross-validation, the initial starting sample is subdivided into approximately K equal size subsamples in a random manner. A single subsample is retained in the K subsamples in order to validate data that will test the model and the left out K-1 subsample will be used as training data. The K subsamples are used exactly once as the validation data right after the cross-validation process is set out to run k times repeatedly. A single estimation is presented after a computation of the K-fold average to give the outcome of K-result. All observations are used for both training and validation and therefore this serve as a primary advantage of this method and will then be used for validation just only once.

Using stratified K-fold cross-validation allows each fold to consist of only similar sizes and proportions of class labels and that mitigates classification problems. The cross-validation procedure is to be repeated n times in repeated cross-validation process to give n random partitions of the original sample. A single estimation is produced when the n results are averaged. For example, we consider K equal to 10 and then we have a dataset that is randomly separated into 10 similar folds of equal sizes generated by K equal that gives us 10-fold CV. Each run consist of 9 subdivisions that are utilized primarily for training set and the last subset is used for validation (see Fig. 3-4). After

each repetition of 10 runs during the procedure, each subset is only used once for validation. After the system has completed running, all the 10 validated results are combined together with the primary purpose of creating an individual of a calculation of out-of-sample.

The K-fold CV obtained a generalization error and K was then wound to 10 in order to balance calculation costs and reliable estimates and to fairly compare with the state of the art.

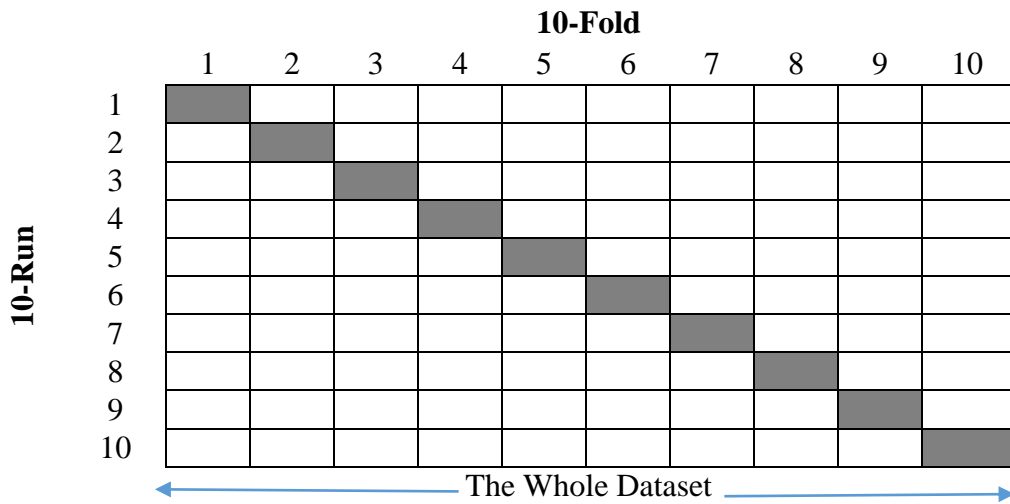


Figure 3.4: 10 fold cross-validation

## Chapter 4

### CLASSIFICATION TECHNIQUES

#### 4.1 Introduction

Classifier is a function or an algorithm that describes how to record substances and their exchanges in the system, that maps every possible input to a finite decision system.

$X$  is input space,  $X = [x_1, \dots, x_d]$   $x \in X$  sample for an input space. We also call  $X$  is feature vector.  $L$  is a finite set of categories to which the input samples belong:  $L = \{c_1, c_2, \dots, c_w\}$ . where  $C$  is the number of classes and  $c_i \in L$  are called labels.

A classifier is like a function  $Y = f(x)$ .  $Y$  is a finite set of decisions, the output set of the classifier. Usually  $Y=L$ , but it can also contain other decisions, such as “no decision”, “reject” (doesn't belong to any category from  $L$ ).

A classifier is function which predict of test data according to train data. Train classifier is a part of main data (training data) that we use for training and test classifier is a test of main data (except data we use for train classifier) which we use for predicting test labels.

In this chapter we study the behavior of three kinds of classifiers namely Support Vector Machine (SVM), K- Nearest Neighbor (KNN), Random Forest (RF) and Decision-Tree.

## 4.2 Support Vector Machine Classifier

Support Vector Machine (SVM) is the most recent classifier in the field of pattern recognition and machine learning. Due to easy implantation and extremely excellent result compared with other classification it is widely used in many applications [16] [35] [36].

The overview of Support Vector Machine (SVM) appeals to researchers from different fields spectrums. The privileges of Support Vector Machines incorporate a great deal of precision and mathematical analysis and evaluation. Below is detailed description with geometric interpretation. Given an A-dimensional N-size training dataset as

$$\{(x_n, y_n) | x_n \in R^A, y_n \in \{-1, +1\}\}, \quad n = 1, \dots, N \quad (4.1)$$

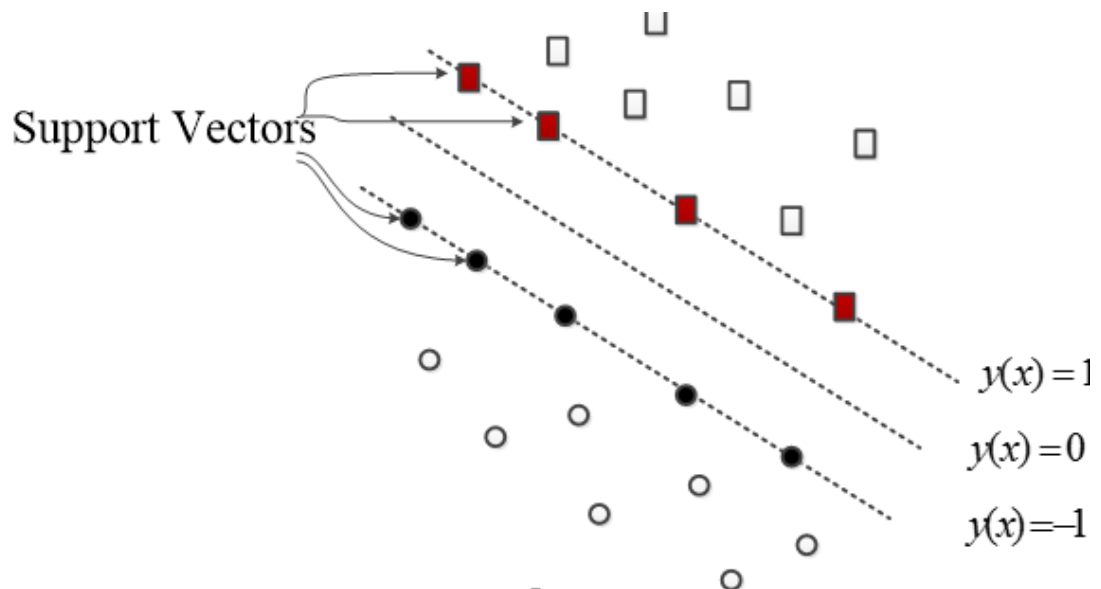


Figure 4.1: Support Vector Machine classifier

where  $y_n$  may be either  $-1$  or  $1$  matching two different classes. Every  $x_n$  is a dimensional vector. The highest margin hyperplane that split two classes is called support vector machine as it is shown in Fig 4.1. Since any hyperplane could be expressed by  $w_x - b = 0$ , whereby  $w$  shows a typical vector towards the hyperplane. We need to choose  $w$  and  $b$  in order to successfully access the margin line that is situated between two parallel hyperplanes. Two parallel hyperplanes are indicating  $w_x - b = \pm 1$ .

### 4.3 K-Nearest Neighbor Classifier

K Nearest Neighbor algorithm is nonparametric technique which is naturally used for classification and regression difficulties. KNN is not a sophisticated classifier. This classifier is one of the simplest classifiers and is utilized in diversity of applications, for example financial anticipating, information compression and genetics.

An ambiguous case can be defined by some feature vectors for classification using a point in the feature space and a KNN classifier separates points in the dataset. For KNN classifier  $K$  is an integer number.

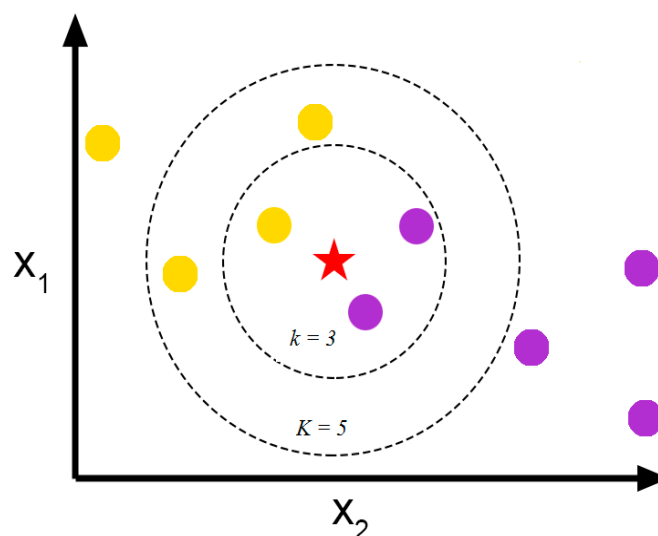


Figure 4.2: K-Nearest Neighbor Classifier

KNN classifier is used in a supervised learning setting. K Nearest Neighbor (KNN) depends on determining the displacement amongst the tests information and every train information to undertake a conclusive decision for presentation (test data). There are different types of distance function such as Murkowsky, correlation, cosine similarity measures and Euclidean distance. Euclidean distance in the KNN algorithm is the commonly used for distance function [37].

To measure the interval between two points for example F and G which are situated in the same phase, several handling operations are used, in which the Euclidean interval function is the most commonly utilized mathematical handling operation. F and G are shown by feature vectors  $F = [x_1, x_2, \dots, x_n]$  and  $G = [y_1, y_2, \dots, y_n]$ , where n is the dimensionality of feature space. To figure out the interval amongst F and G, the standardized Euclidean metric is generally used by

$$\text{Int}(F,G) = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (4.2)$$

#### **4.4 Decision Trees Classifier**

In a case whereby all the variables in our predictor are continuous, to exercise classification we need to divide our training data into two parts by choosing a single variable that will be segregated into two halves through the rule of partitioning. The segregation will continue until a barricading parameter is reached.

Hyper rectangles also known as leaves or terminal nodes are used to partition predictor space for the above mention procedure to take place. The training data will then be converted and reviewed as test data points and that gives way to a class classification with the highest number of training data points in the region. Assessing the magnitude



of training points from every class in the region permits a high yield of probability estimates [38] [39].

One viable and possible option amongst many to use the splitting rule is to use a split that provides the highest decrease in deviance tree defined as

$$X = \sum_i X_i \text{ \& } X_i = -2 \sum_k n_{ik} \log p_{ik} \quad (4.3)$$

where  $n_{ik}$  represents the sum of observation from Class X the number of observation from Class k in the  $i_{th}$  terminal node and  $p_{ik}$  is the likelihood of an observation in the  $i_{th}$  terminal node being from Class X. When the drop in error rate or deviance drops below a threshold value the partitioning stops.

In a case the response is continuous, the possibility of building a regression tree increases. Although trimming the tree to reduce the number of terminal nodes is a viable option to consider, however it tends to have the effect of reducing variability and risk increasing bias [39].

## Chapter 5

# DETECTION OF ALZHEIMER DISEASE BASED ON KEY SLICES AND DATA FUSION

### 5.1 Introduction

For this chapter a clear description of detecting AD from HC subjects will be fully outlined. In the first section of this thesis we gave a descriptive outline regarding the AD in detail and also supported our discussion with preprocessing data downloaded from OASIS database containing cross-section of MRI data, longitudinal MRI data and additional data for each subject. Since the calculation region of interest for the WB is result depriving, a KS selection that will serve as the wholly scanner of the brain was introduced. The preprocessing section include region detection to section region of interest (ROI) based on displacement field (DF) method. In the next section, we call feature extraction we extract energy, contrast and homogeneity metrics along with feature vectors generated by PCA and probability distribution function (PDF) methods for each key-slice. In the last section features coming from each key-slice are combined through feature fusion for improved accuracy. In all implementation of propose method was drawn Fig 5.1.

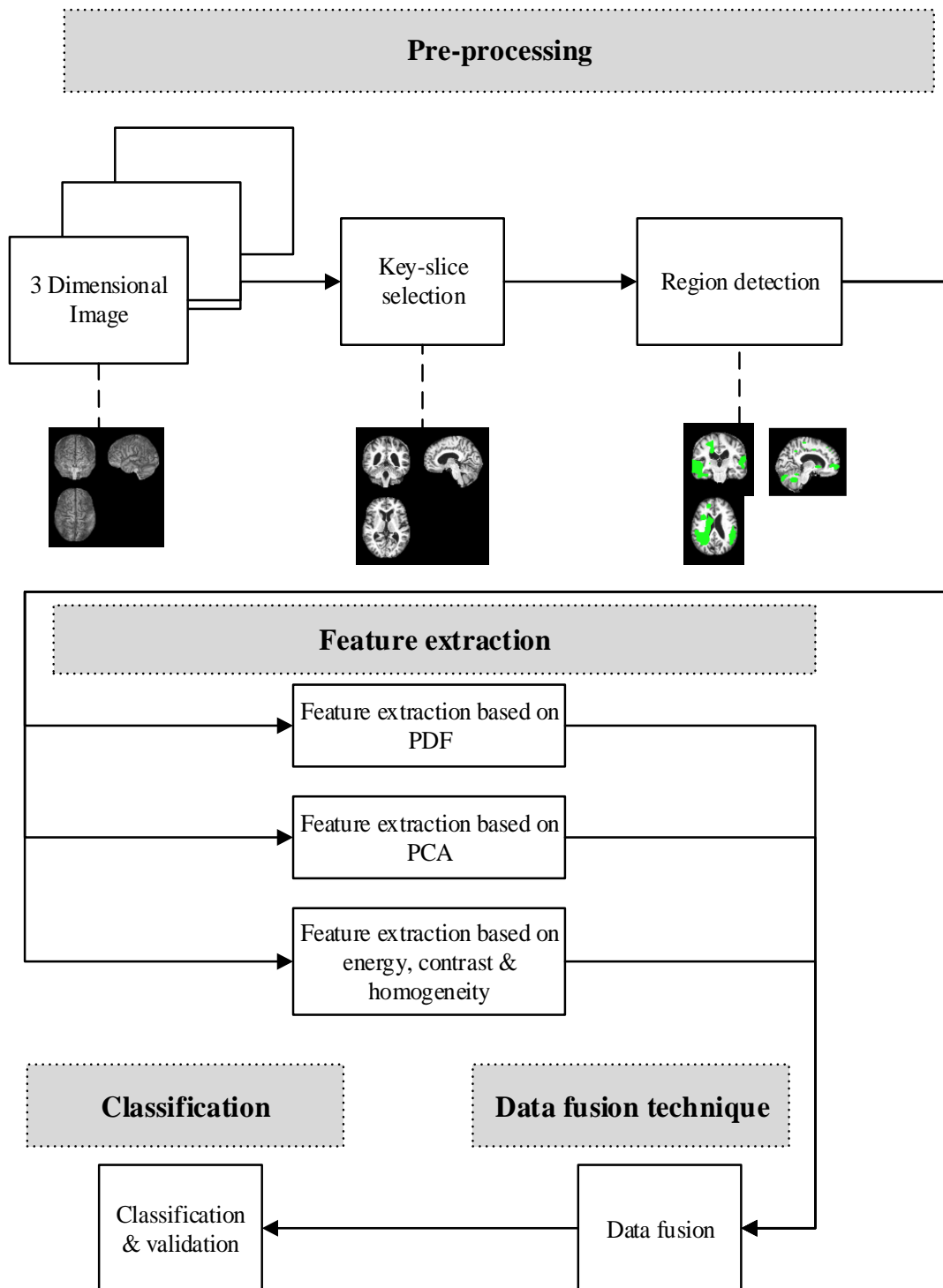


Figure 5.1: Block diagram of proposed AD detection method.

## 5.2 Key-Slice Selection

A KS selection process which choose KSs having structures indicating AD from HC was introduced since the computation of displacement field on the wholly scanner of the brain was time-consuming [40].

A method based on fisher criterion  $J(w)$  to select the best KS was presented as follows, given in Equation (5.1):

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \quad 5.1$$

In equation 5.2  $S_B$  is indicative of scatter matrix for classes L1 and L2 also in equation 5.3  $S_W$  is indicative within class scatter matrix for classes L1 and L2 defined below as

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \quad 5.2$$

$$S_W = \sum_{x_i \in c_1} (x_i - \mu_{L1})(x_i - \mu_{L1})^T + \sum_{x_i \in c_2} (x_i - \mu_{L2})(x_i - \mu_{L2})^T \quad 5.3$$

Where  $w = S_W^{-1}(\mu_{L1} - \mu_{L2})$  and  $\mu_{li}$  is the mean of data in each class.

## 5.3 Region Detection

local and global registrations as well as the registration approximations and the rigid elements are contained within the shape registration, which were attained ‘through both processes of brain-masking and co-registration. The local registration finds The “displacement field (DF)” which is located by the local registration between the reference and moving images (see Fig. 5.2).

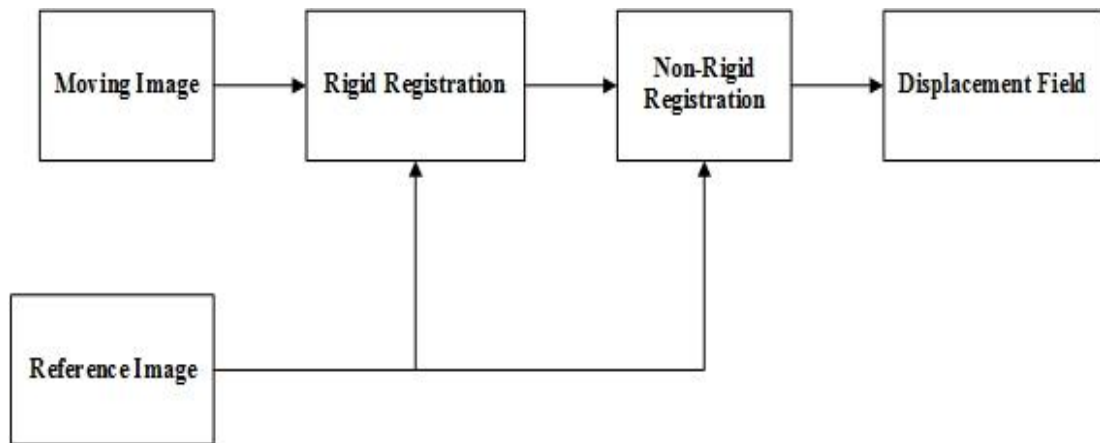


Figure 5.2: The diagram of displacement field generation.

The rigid registration is a necessary and imperative preprocessing procedure since it has the ability to fix deformation resulting from positioning, movement, and posture of patients. The non-rigid registration reflecting the shape and distortion of the diseases can be thought of as the motion valuation task among a HC brain (moving) and an AD brain (reference). Numerous kinds of solutions existing to give a clear and concise descriptive discussion of this task are spline function based techniques, fluid techniques, optical-flow techniques, phase-correlation techniques, elastic techniques to list a few. An example of the first above listed solution could be regarded as parametric, since, that sort of explanation requires to mitigate related spline-based function criterions.

The second kind is very challenging to determine in terms of finding the local search range and also finding the global optimal points because it requires a series and network of computational resources. Three other methods can be regarded as non-parametric because they discover the DF by deciphering a predefined tangible type with the usage of Partial Differential Equation (PDE).

The new technique, the level-set motion, that is created on the foundation of a level-set evolution theory render a moving-image ( $I_1$ ) to transform along its pitch direction till it comes to a halt to reference-image ( $I_2$ ). The DF can be written as

$$\frac{dV}{dt} = (I_2 - I_1(V)) \frac{\nabla I_1(V)}{|\nabla I_1(V)|} \quad 5.4$$

The equation 5.4  $V$  displays displacement fields that are solved by repetition of algorithms described in the reference.

## **5.4 Feature Extraction**

A pixel is separated through the enactment of a feature extraction technique that relies solely on DF analysis application. The regions of decreased gray matter size obtained using The DF analysis obtains intrinsic images of regions of deformed portions of the brain that have a drop in the overall spread of the gray matter size. Affected patients are examined using a 2D mask that measures the density pixels of the deformation part of the brain and then extract pixels and then perform feature extraction based on PDF, extraction based on PCA and extraction based on contrast, homogeneity and energy list all the sample figures as a feature vector. Obtained dataset is sampled into 10 folds with similar fold length of AD and HC samples in each fold.

### **5.4.1 Feature Extraction Based on PDF**

A PDF can be described as a vector of probabilities representing the probability of the pixel values and also defined as a raw feature vector extracted from image that belong into various disjointed intervals. It could also be defined as a statistical description of the distribution of occurrence probabilities of pixel values that can be considered a

feature vector, defined as bins [41] [12]. The PDF extracted from voxels or pixels can be calculated as follows:

$$H = [p_1, p_2, \dots, p_m], p_i = \frac{\eta_i}{N}, i = 1, 2, \dots, m \quad 5.5$$

$\eta_i$  represents the total number of pixels inside the  $i_{th}$  bin,  $m$  represents total bins, and  $N$  represents the sum of pixels in the 2D mask. In this thesis, we experimented having to consider optimal number of bins to be 20 after having indulged in a trial and error method of number of bins.

#### 5.4.2 Feature Extraction Based on Statistical Parameters

Consider  $V$  represent 2 dimensional MRI data for each key-slice and  $V(x,y)$  as pixel located at  $(x,y)$ . we utilized Energy  $E$ , Contrast  $C$ , Homogeneity  $H$  as the representation data  $V$ . They're expressed as:

$$E = \sum_{i,j} V(x,y)^2 \quad 5.6$$

$$C = \sum_{i,j} |i-j|^2 V(i,j) \quad 5.7$$

$$H = \sum_{i,j} \frac{V(i,j)}{1+|i-j|} \quad 5.8$$

The triplet (energy, contrast, homogeneity) extracted from all key-slice of brain and then queued into a row vector.

### 5.4.3 Feature Extraction Based on PCA

Eigen-brain is a mathematical procedure that utilizes orthogonal advancement facilitated by PCA to match possibly corresponding and correlated variables into a set of uncorrelated values that could be defined as principal components. Therefore, this means 2D images are elevated naturally to become 2D eigen-brains.

In a case size  $M \times N$  is derived from normalized dataset  $X$ , assuming  $M$  represents number of samples and  $N$  represents number of features. Normalized  $Z$  represents centered and scaled unit variance is derived by subtracting the mean value and dividing the resulting difference by its standard deviation value and thus this is represented by the formula below

$$Z = \frac{X - \mu(X)}{\sigma(X)} \quad 5.9$$

The covariance of matrix  $C$  with size of  $N \times N$  is expressed in the formula below

$$C = \frac{1}{M - 1} Z Z^T \quad 5.10$$

$M - 1$  instead of  $M$  is used to represent variance approximation value

Third, we express the Eigen decomposition of  $C$ :

$$C = U \Lambda U^{-1} \quad 5.11$$



U is represented by  $N \times (M - 1)$  matrix, whose confinements are the eigenvectors of matrix C covariance, and matrix  $\Lambda$  is represented by  $(M - 1) \times (M - 1)$  diagonal matrix whose diagonal values are eigenvalues of C, each correlating to an eigenvector N. It is usual process to sort the eigenvalue matrix  $\Lambda$  and matrix U by reducing eigenvalue  $u_1 > u_2 > \dots > u_N$ . To view the *i*th Eigen-brain  $u(i)$ , the *i*th column of U was reshaped to an image.

In this thesis we selected 10 first eigenvalues from each key-slice of brain.

## 5.5 Data Fusion

This part introduces data fusion technique to improve the accuracy of the proposed AD classification method. The goal of the data fusion technique is to combine the data from two or more distinct multiple source (vectors, classifiers) to improve performance [42].

In the scheme of source data fusion, the top features selected are based on our approach according to scheme of the source data fusion with an affirmative description, from selective images knotted into a one feature vector. described in the last chapter, from variety Images, are concatenated into a single feature vector. In a case  $fV_1, fV_2, \dots, fV_n$  are featured feature vectors created using presented feature selection methods for each cluster, the feature vector fusion (FVF) is then:

$$fVf = [fV_1, fV_2, \dots, fV_n]_{1 \times \sum_{i=1}^n m_i} \quad 5.12$$

whereby  $m_i$  represents vector length for  $fV_i$ . This innovated feature vector is then used for classifying procedures for feature contraction [43].

## **5.6 Classification**

The final step for distinguish AD from HC subjects is called Classification which evaluate performance serve as the last stage for distinguishing AD. In this thesis few classifiers have been introduced for AD classifications including support vector machine (SVM) classifier, KNN and decision tree (DT) classifier.

## Chapter 6

### RESULTS AND DISCUSSIONS

#### 6.1 Key Slice Selection

The feature extraction on the WB was time consuming, and therefore we decided to select the best layer of brain in axial direction, sagittal direction and coronal direction based on fisher criteria rule. The optimal number of slices are determined by fisher criteria. The fisher criteria between AD and HC group is calculated for all key slices and then selected as number of top discriminative key slices.

In this study Fisher Criteria aid us to find the best key slices with the most discriminative pixels for the feature extraction and classification process. To find Fisher Criteria coefficient for each axis, we calculated the coefficient of fisher criteria for each slice. If the value of Fisher Criteria is equal to zero for each slice, the pixels of that layer doesn't have any difference between AD and HC, and the fact that the value of Fisher Criteria is high shows that these layers AD and HC differ from each other.

The diagram for Fisher Criteria for coronal axis is shown below in figure 6.1. For selection key slices in this axis, we choose an area that is higher than half the maximum value. As it's displayed in the diagram below, the area between 60 to 155 has more discriminative pixels compared to other slices. Since we want to have uncorrelated

slices, we have to pick ten key slices from 60 to 150 with increasing step of 10. Figure 6.4 below shows key slices for coronal axis.

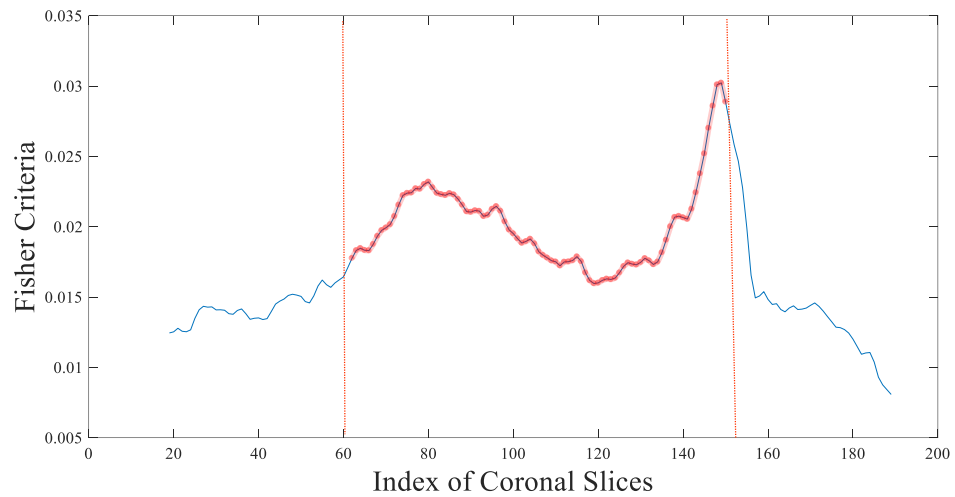


Figure 6.1: Curve of Fisher Criteria for coronal direction

The diagram of fisher criteria for sagittal axis is displayed in figure 6.2. For selection key slices in this axis we chose an area that is higher than half of the maximum value. As it's displayed in the diagram the area between 40 to 130 has more discriminative pixels compare with other slices. Since we want to have uncorrelated slices, we have to pick ten key slices from 40 to 130 with increasing step of 10. Figure 6.5 shows key slices for sagittal axis.

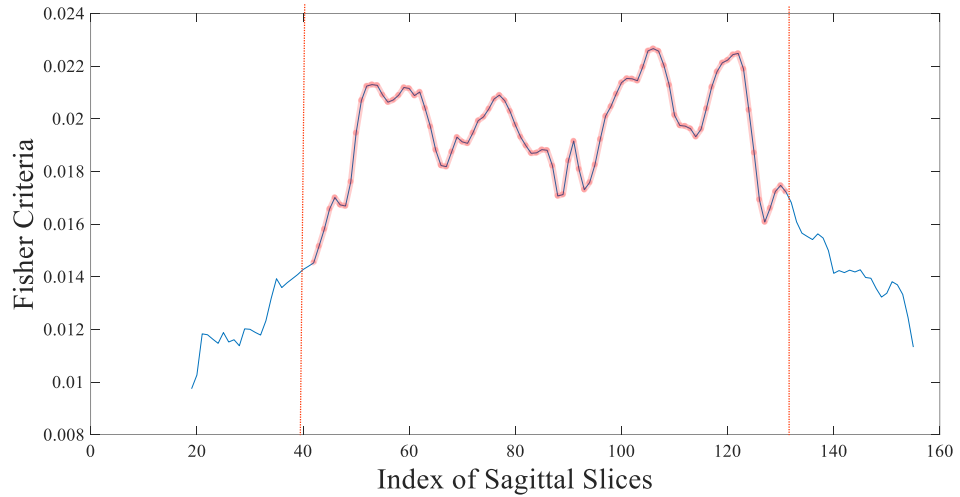


Figure 6.2: Curve of Fisher Criteria for sagittal direction

The diagram of fisher criteria for axial axis is shown in figure 6.3. For selection key slices in this axis we chose an area that is higher than half the maximum value. In the diagram below the area between 60 to 100 has more discriminative pixels compared to other slices. Since we want to have uncorrelated slices we pick ten key slices from 60 to 100 with increasing step of 10. Figure 6.6 shows key slices for coronal axis.

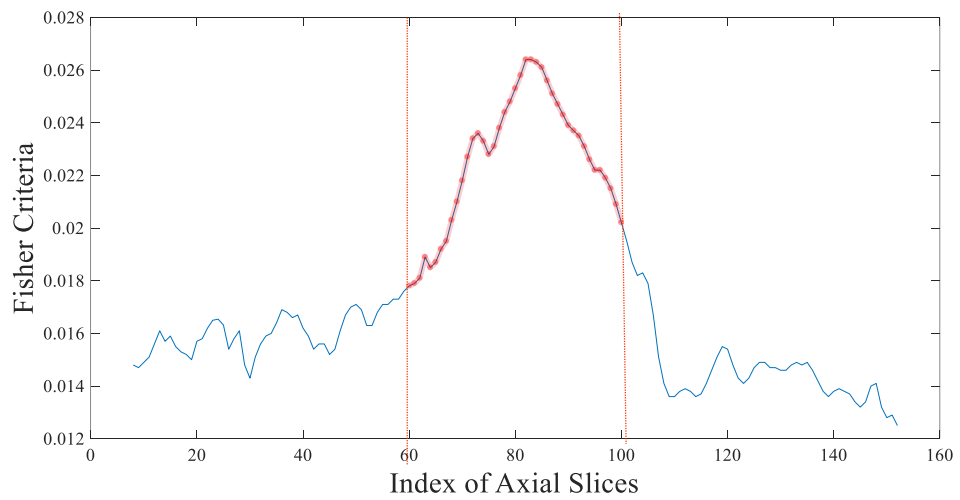
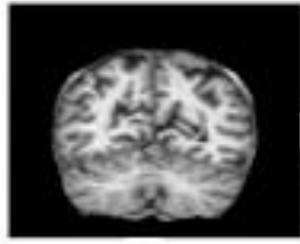
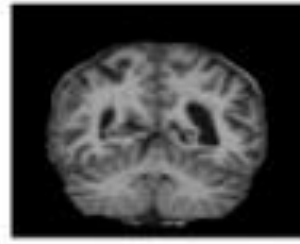


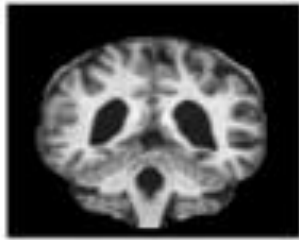
Figure 6.3: Curve of fisher criteria for axial direction



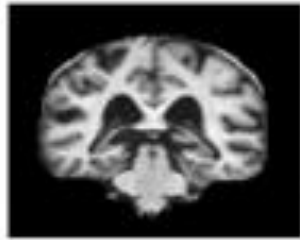
Slice # 60



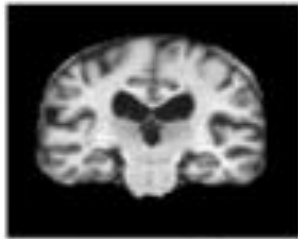
Slice # 70



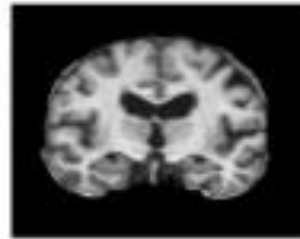
Slice # 80



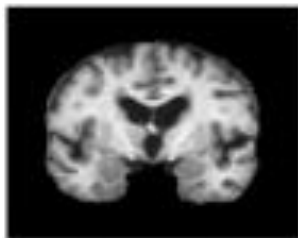
Slice # 90



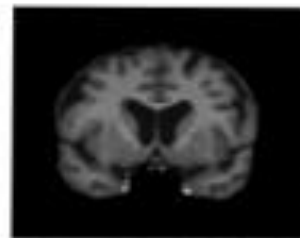
Slice # 100



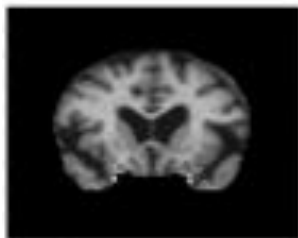
Slice # 110



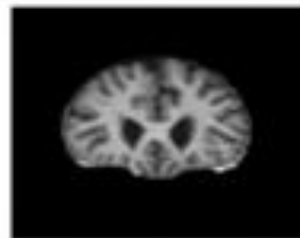
Slice # 120



Slice # 130



Slice # 140



Slice # 150

Figure 6.4: Key-Slice of coronal from 60 to 150 with up-sampling 10

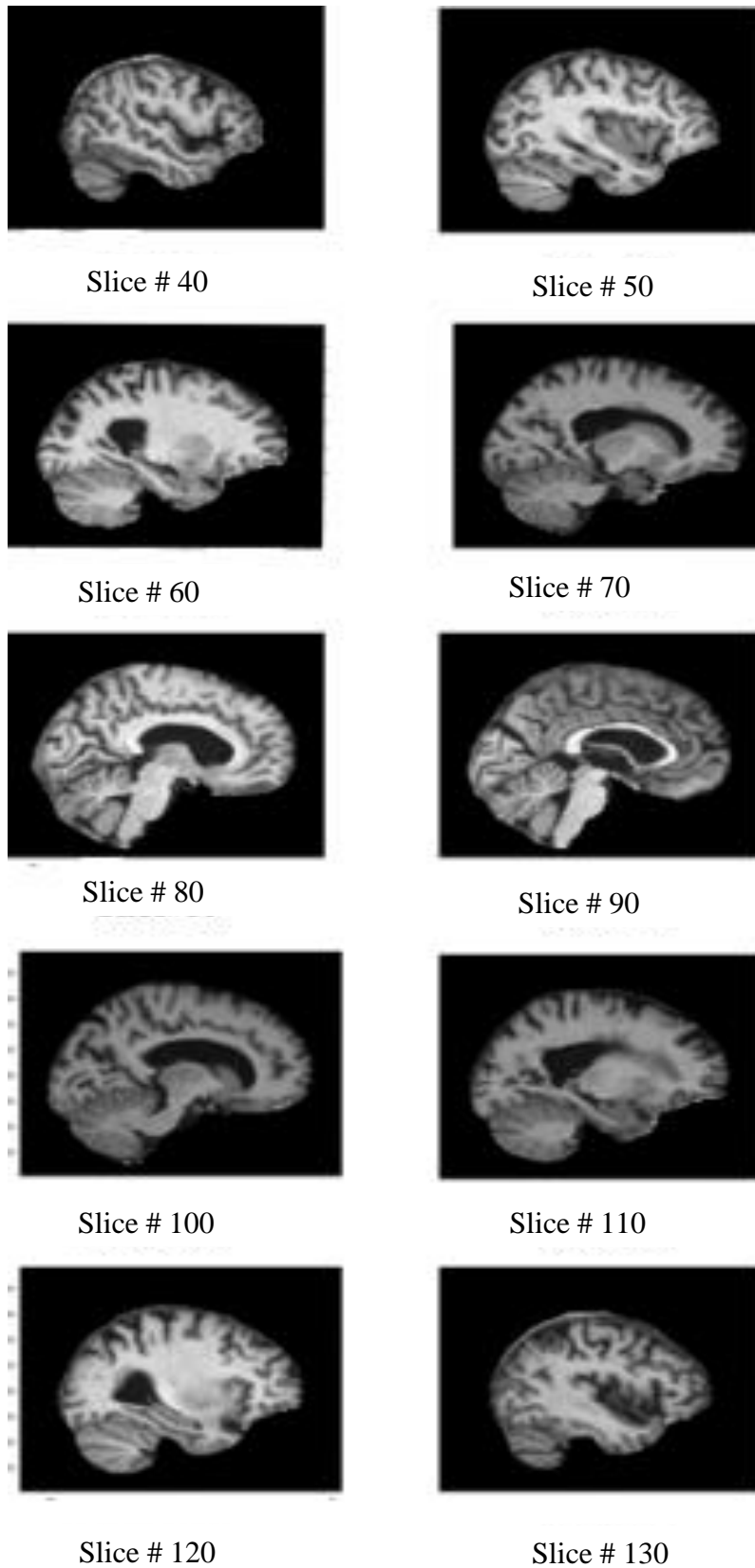
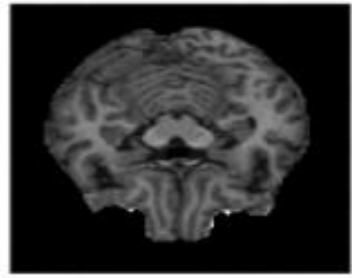
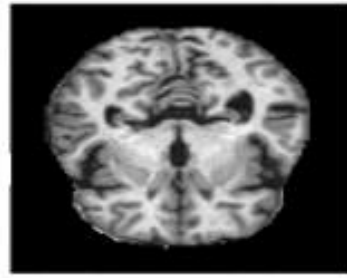


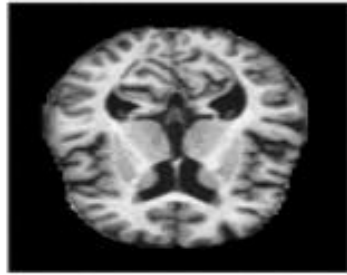
Figure6.5:Key-Slice of sagittal from 40 to 130 with up-sampling 10



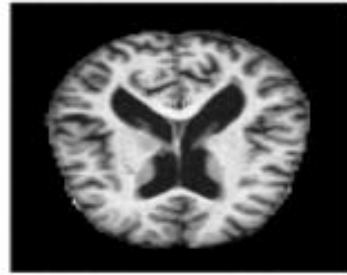
Slice # 60



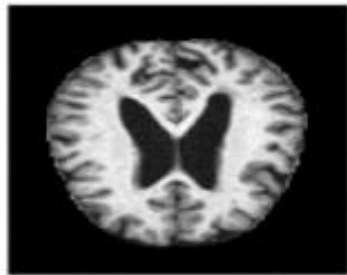
Slice # 70



Slice # 80



Slice # 90



Slice # 100

Figure 6.6: Key-Slice of axial from 60 to 100 with up-sampling 10

## 6.2 Region Detection

Firstly, we randomly selected one AD and one HC from our original data and we obtain DF for each key slice. The same process is repeated for all AD and HC samples in the train dataset (27 AD and 87 HC). This process is repeated because we have different regions for every sample from the dataset. We use all the images to obtain an average for every image. For the next step, we define threshold to generate the deformation map(region) in the brain.



$$ROI = \{(x, y) | D(x, y) | > T\} \quad (6.1)$$

Here  $D(x, y)$  represents the displacement field at the point of  $(x, y)$ ,  $|\cdot|$  represents the magnitude and  $T$  represent the threshold. In other words, we consider the point with magnitude larger than  $T$ . According to the experiment we consider value of  $T$  as 5. Smaller value of  $T$  may indicate more noises in the computed displacement field, whereas with large number of threshold we will lose realistic deformation.

The process of ROI (Region of Interest) for each axis is shown by flowchart Fig 6.7. In our first step we calculate number of iterations. The number of iteration is equal to number of subject of ADs times number of subjects of HC. We then randomly selected one AD and one HC then we calculated DF for all the subjects and then we calculated the mean subjects in the last step according to the threshold we defined as ROI. In figure 6.8 we show ROI for key-slices in axial axis with green point. In figure 6.9 we show ROI for key-slices in coronal axis with green point and also in figure 6.9 we show ROI for key-slices in sagittal axis with green point.

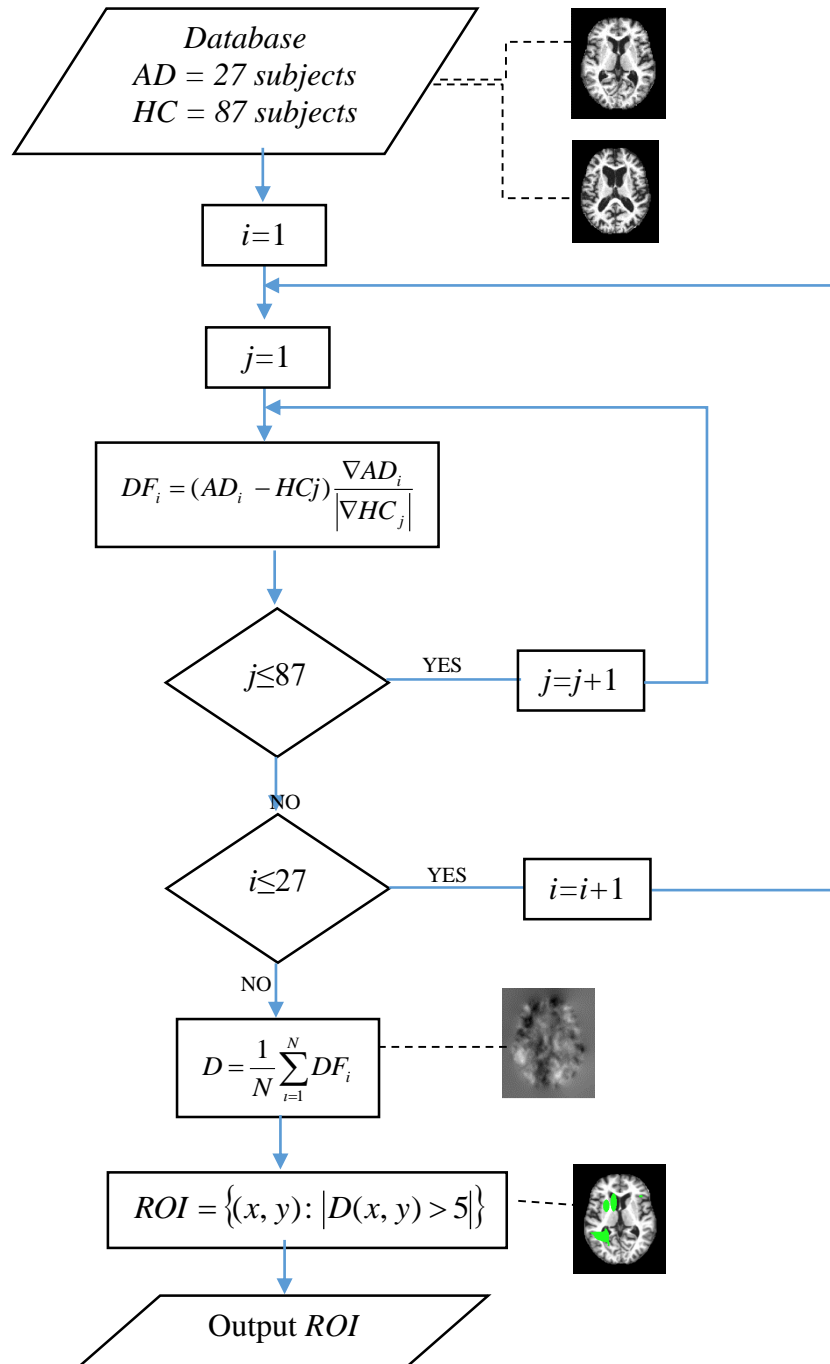
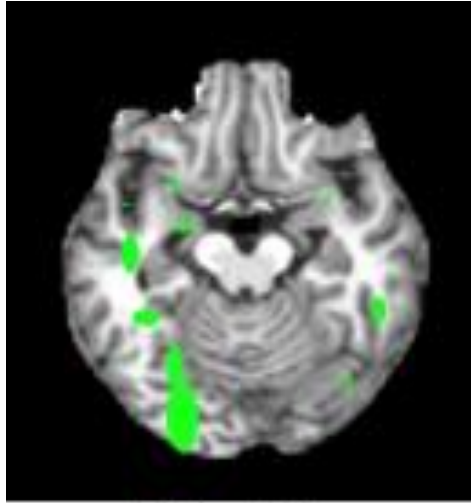
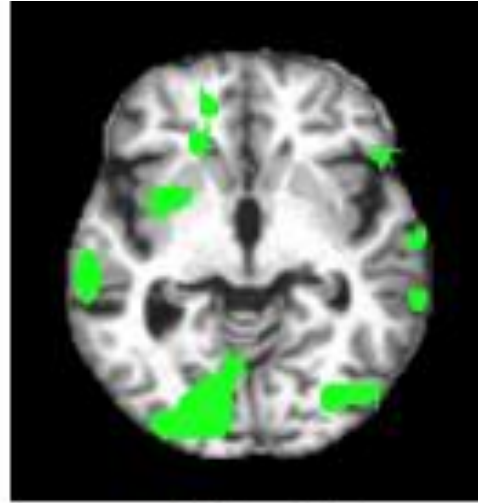


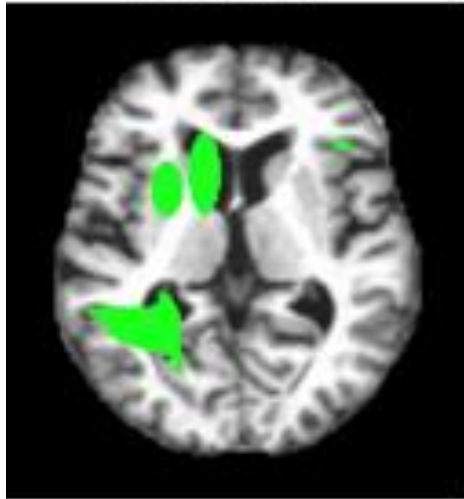
Figure 6.7: Flowchart of the proposed Region Detection Method



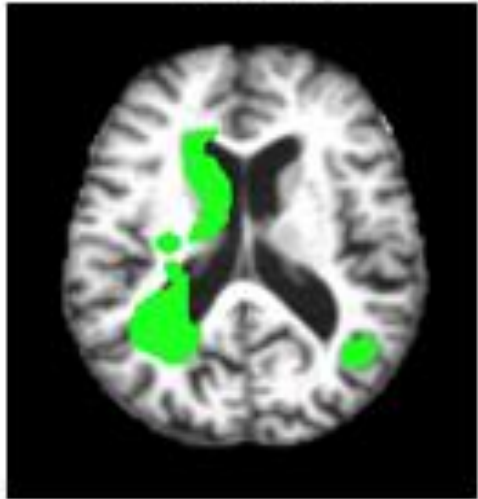
Slice # 60



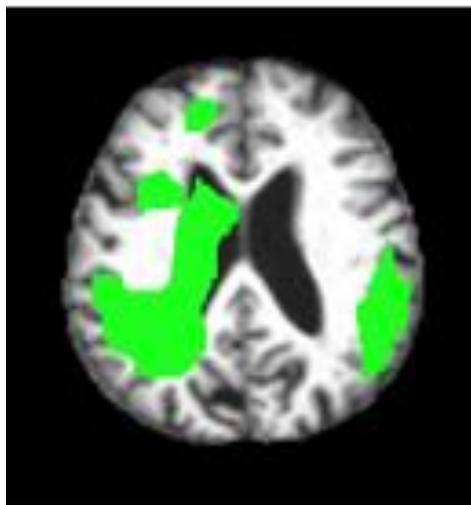
Slice # 70



Slice # 80



Slice # 90



Slice # 100

Figure 6.8: Brain Mask of Axial View from 60 to 100 with Up-Sampling 10

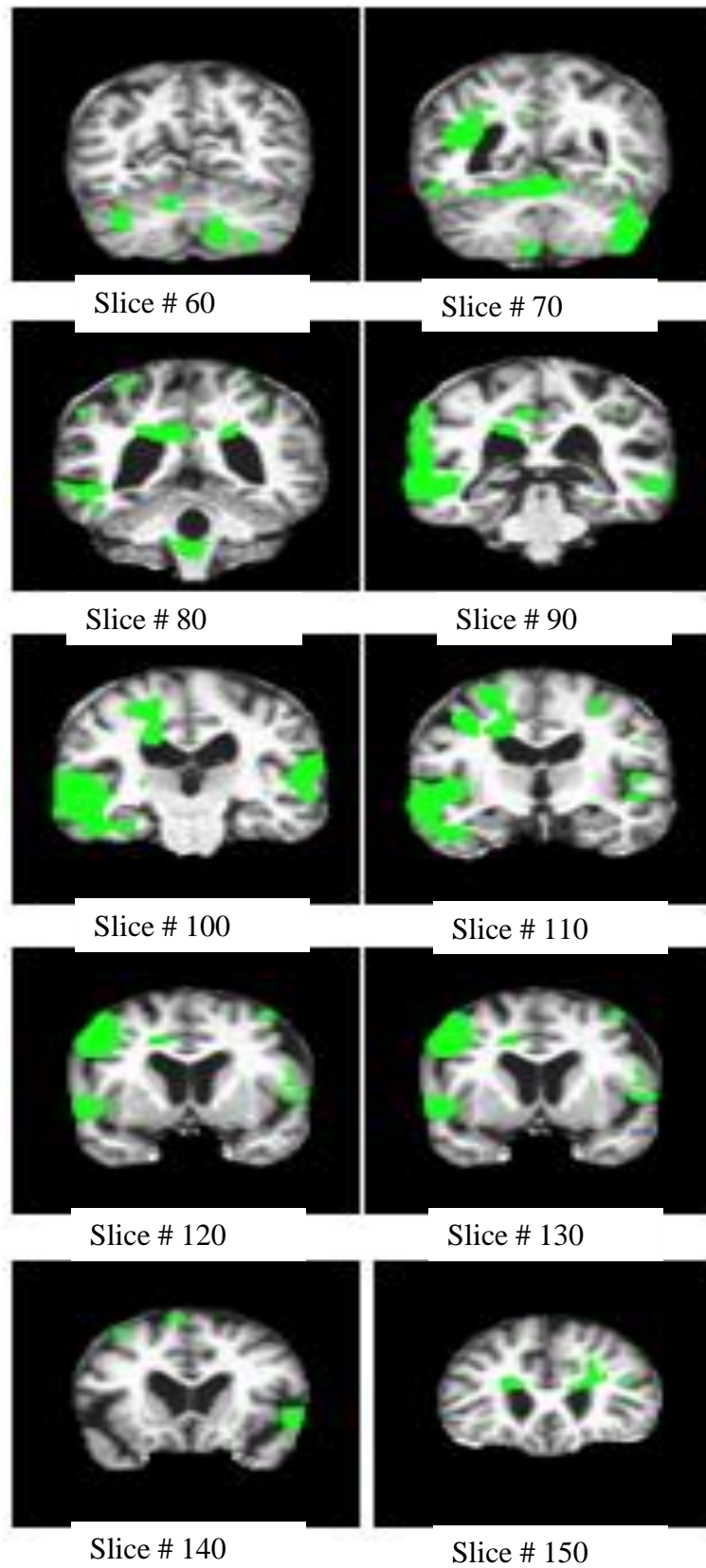


Figure 6.9: Brain Mask of Coronal View From 60 to 150 with Up-Sampling 10

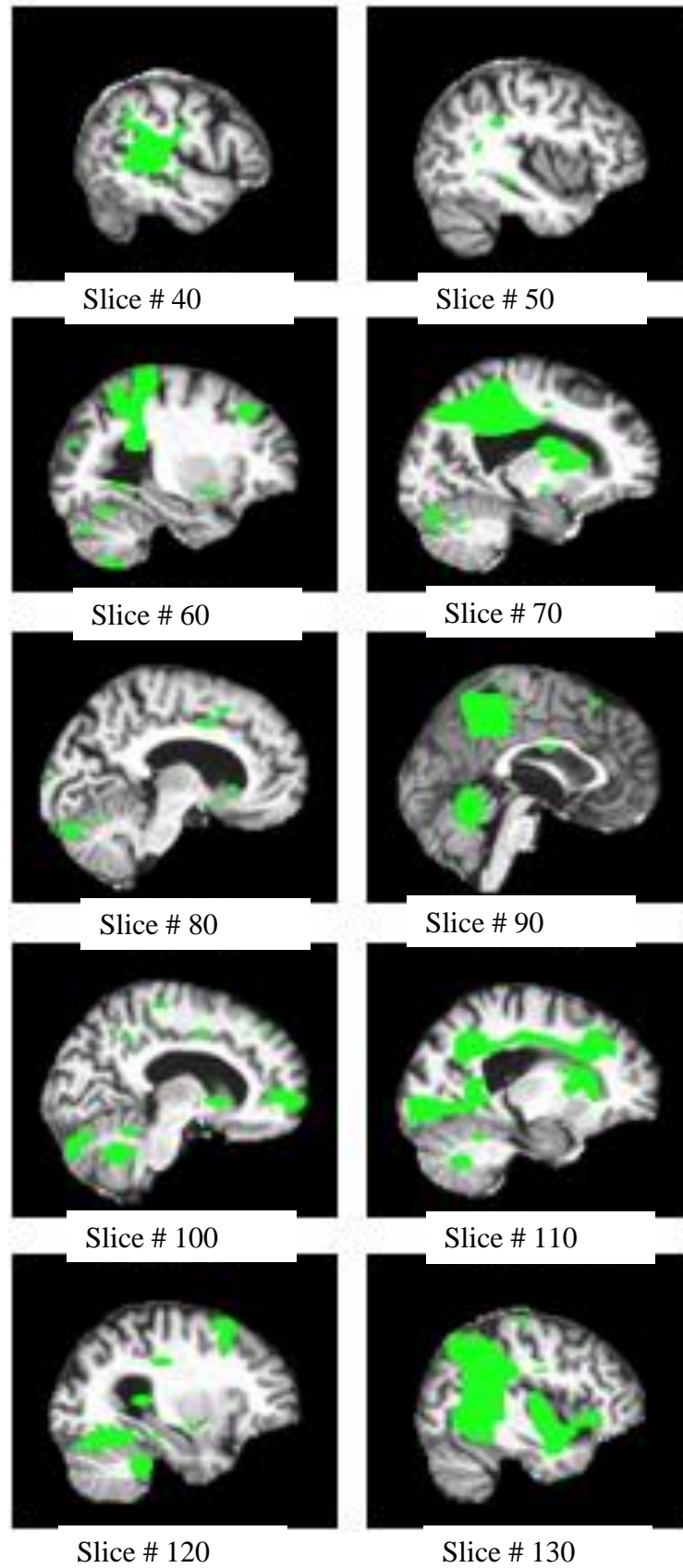


Figure 6.10: Brain Mask of Sagittal View From 60 to 150 with Up-Sampling 10

### **6.3 Feature Extraction and Classification Comparison**

In this section, we report accuracy of the 10-fold cross validation for detection of AD calculated over the OASIS data. In each axis for each slice we perform feature extraction based on PCA (10 features), PDF (20 features), Energy, Contrast and Homogeneity (ECH - 3 features). For each slice 33 features were extracted. Coronal axis, sagittal axis and axial axis have 10 slices, 10 slices and 5 slices respectively. This means that 330 features are extracted from coronal and sagittal axis and 165 features are extracted from axial axis.

The experimental results show that the proposed methods are comparable with classical methods introduced in the literature. Table 6.1 shows the accuracy (%) of proposed method without using any masks. In this regard, we investigated the accuracy of different feature extraction techniques for each individual axis as Coronal, Sagittal, and Axial axis. The accuracy of extracted feature evaluated by utilizing different classifiers. It is clear that the PCA has the highest performance among all axis for KNN and SVM classifiers. The SVM classifier has the best performances when it applied to all feature extraction method in all axis. Furthermore, we used all three axes to evaluate the performance. As it is shown in Table 6-1 the accuracy is improved by having all three axes. Further improvement can be achieved if we fuse all axes. The highest accuracy is belonging to PCA using linear SVM with accuracy of 83.1%. The same approach can be done by utilizing mask as it is illustrated in Table 6-2. The highest accuracy is achieved when we applied fusion techniques followed by SVM classifier and it reaches to 88.1%.

Table 6.1: The accuracy of the proposed method without using mask

Axis	Feature extraction methods	classifier		
		Decision Tree	Linear SVM	KNN
Coronal axis	PCA	67.7	82.7	80.3
	PDF	68.5	75.6	72.4
	Energy, Contrast Homogeneity	73.2	78.0	78.0
Sagittal axis	PCA	73.2	82.1	81.1
	PDF	75.6	78.0	71.7
	Energy, Contrast Homogeneity	75.6	81.1	71.7
Axial axis	PCA	66.9	82.5	81.0
	PDF axial	73.2	79.5	77.2
	Energy, Contrast Homogeneity	77.2	79.5	70.1
Three axis	PCA <sub>coronal</sub> + PCA <sub>sagittal</sub> +PCA <sub>axial</sub>	71.7	<b>83.1</b>	<b>82.7</b>
	PDF <sub>coronal</sub> + PDF <sub>sagittal</sub> +PDF <sub>axial</sub>	69.3	76.4	74.8
	ECH <sub>coronal</sub> + ECH <sub>sagittal</sub> +ECH <sub>axial</sub>	70.9	77.2	74.8
PCA <sub>all</sub> + PDF <sub>all</sub> + ECH <sub>all</sub>		<b>78.7</b>	82.7	80.3

Table 6.2: The accuracy of the proposed method with mask

Axis	Feature extraction methods	classifier		
		Decision Tree	Linear SVM	KNN
Coronal axis	PCA	76.4	85	80.3
	PDF	63.0	78.7	79.5
	Energy, Contrast Homogeneity	72.4	80.3	79.5
Sagittal axis	PCA	63	81.1	81.1
	PDF	70.1	84.3	77.2
	Energy, Contrast Homogeneity	69.3	79.5	76.4
Axial axis	PCA	64.6	73.2	75.6
	PDF	70.1	79.5	74.8
	Energy, Contrast Homogeneity	71.7	76.4	78
Three axis	PCA <sub>coronal</sub> + PCA <sub>sagittal</sub> +PCA <sub>axial</sub>	<b>80.3</b>	80.3	83.5
	PDF <sub>coronal</sub> + PDF <sub>sagittal</sub> +PDF <sub>axial</sub>	72.4	85	78
	ECH <sub>coronal</sub> + ECH <sub>sagittal</sub> +ECH <sub>axial</sub>	68.5	82.7	79.5
Proposed method (PCA <sub>all</sub> + PDF <sub>all</sub> + ECH <sub>all</sub> )		78.7	<b>88.1</b>	<b>85.1</b>

In Table 6-3, we compared proposed method ( $PCA_{all} + PDF_{all} + ECH_{all} + LSVM$ ) with other papers. Accuracy of the proposed method is higher than other methods, as well as, the sensitivity and specificity is comparable with the state of the art. The database used in this work, contains brain MRI scans of 30 AD and 97 HC subjects. The database adopted in this thesis can be considered to be unbalanced by means of equal number of AD and HC samples. We had to use an unbalanced database due to difficulties faced to access private balanced AD databased. Despite of this disadvantage, the accuracy, sensitivity and specificity performances of the proposed methods are either higher or comparable with other methods. We can't directly compare them because the databases are different but we can claim that having a balanced database would contain equal number of samples in both classes, hence, the proposed method may be better analyzed and compared with the alternative methods in the literature.

Table 6.3: Comparison Our Method with Other Methods

Method (Author)	Source of data	Subjects (AD/HC)	ACC (%)	SEN (%)	SPE (%)
MSD + RBF SVM (Papakostas et al., 2015) [44]	Private	49/49	85.00	78.00	92.00
VV + RBF-AB-SVM (Savio et al., 2011) [45]	Private	49/49	86.00	80.00	92.00
GM+WM+SVM(Khedher et al., 2015) [46]	ADNI	188/229	88.49	90.39	86.17
EB+WTT+RBF-KSVM (Dong et al., 2015) [17]	OASIS	28/98	86.71	85.71	86.99
BRC + IG + VFI (Plant et al., 2010 ) [32]	Private	32/18	78.00	65.63	100
<b>Proposed method (<math>PCA_{all} + PDF_{all} + ECH_{all} + LSVM</math>)</b>	OASIS	30/97	88.10	80.76	91.08



## Chapter 7

### CONCLUSION AND FUTURE WORK

#### 7.1 Conclusion

In this study, our focus was to propose new approaches to improve the accuracy of detection of Alzheimer's disease. In this regards, firstly we used key-slice in 3D to reduce the dimensionality of MRI images. In the previous studies key-slice is used in 1-D only. Next new feature extraction method is proposed by utilizing 2D slices in three orthogonal directions. The proposed method includes the calculation of Fisher Criterion between the AD and HC groups in order to select key-slices in the coronal, sagittal and axial directions. Experimental results show the effectiveness of the proposed feature extraction with respect to the state-of-the-art. Finally, extracted features coming from each key-slice are combined through feature fusion for improved accuracy. Experimental results show fusion method that used with brain mask generates higher or comparable results compared with other feature extraction techniques in the literature. Furthermore, we adopted three individual classifiers to compare our methods and the results show that SVM outperforms DT and K-NN classifiers. Finally, the proposed method achieved higher or comparable sensitivity, specificity and accuracy scores for the detection of AD among the other methods available in the literature.

## **7.2 Future Work**

Future study will focus on many aspects, such as data fusion technique including decision fusion which involves majority voting and also we will add more AD in our study. Feature selection process is another approach for selecting more discriminative feature such as genetic algorithm, which will possibly help an increase in classification accuracy. Finally, deep learning algorithms such as Convolutional Neural Networks (CNN) will be adopted to increase performance in classification process for the MRI images for further improvement in classification rate.

## REFERENCES

- [1] R. J. Castellani, G. Perry, G. L. Iverson, "Chronic Effects of Mild Neurotrauma: Putting the Cart Before the Horse?," *Journal of Neuropathology & Experimental Neurology*, vol. 74, no. 6, p. 493–499, 2015.
  
- [2] Y. Zhang, J. Yang, J. Yang, A. Liu and P. Sun, "A Novel Compressed Sensing Method for Magnetic Resonance Imaging: Exponential Wavelet Iterative Shrinkage-Thresholding Algorithm with Random Shift," *International Journal of Biomedical Imaging*, vol. 2016, p. 10, 23 Feb 2016.
  
- [3] R.Castellani, G.Perry , "The complexities of the pathology-pathogenesis relationship in Alzheimer disease.," *Elsevier*, vol. 88, no. 4, pp. 671-676, 2014.
  
- [4] K. Hahn, N. Myers, S. Prigarin, K. Rodenacker, A. Kurz, H. Förstl, C. Zimmer, A.M. Wohlschläger and C. Sorg, "Selectively and progressively disrupted structural connectivity of functional brain networks in Alzheimer's disease Revealed by a novel framework to analyze edge distributions of networks detecting disruptions with strong statistical evidence," *NeuroImage*, vol. 81, pp. 96-109, 2013.

- [5] S. Goh, Z. Dong and Y. Zhang, "Mitochondrial Dysfunction as a Neurobiological Subtype of Autism Spectrum Disorder," *JAMA Psychiatry*, vol. 71, pp. 665-671, 2014.
- [6] M. Murphy, J. Huston, C. Jack, K. Glaser, A. Manduca, J. Felmlee and R. Ehman, "Decreased brain stiffness in Alzheimer's disease determined by magnetic resonance elastography," *J Magn Reson Imaging*, vol. 34, no. 3, pp. 494-498, 2011.
- [7] R. Brookmeyer, E. Johnson, K. Ziegler-Graham and H. Arrighi, "Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement*," *Alzheimer's & Dementia*, vol. 3, no. 3, pp. 186-191, 2007.
- [8] M. Kantanen, S. Kiuru-Enari, O. Salonen, M. Kaipainen and L. Hokkanen, "Kantanen Subtle neuropsychiatric and neurocognitive changes in hereditary gelsolin amyloidosis (AGel amyloidosis)," *PeerJ*, vol. 2, p. e493, 2014.
- [9] Y. Song and J. Wang, "Overview of Chinese research on senile dementia in mainland China," *Ageing Research Review*, vol. 9, pp. S6-S12, 2010.
- [10] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji and J. Yang, "Exponential wavelet iterative shrinkage thresholding algorithm for compressed sensing magnetic resonance imaging," *Information Sciences*, vol. 322, p. 115-132, 2015.

- [11] Y. Zhang, S. Wang, G. Ji and Z. Dong., "Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 10, no. 1, pp. 116-117, 2015.
- [12] I. Beheshti and H. Demirel, "Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease," *Comput. Biol. Med.*, vol. 64, p. 208–216, 2015.
- [13] D. S. Z. X. B. K. S. M. R. a. C. D. Z. Lao, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," *Neuroimage*, vol. 21, no. 1, p. 46–57, 2004.
- [14] Y. Fan, D. Shen, and C. Davatzikos, "Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM," *Med. Image Comput. Comput. Assist. Interv.*, vol. 8, no. Pt 1, p. 1–8, 2005.
- [15] Z. Dong, Y. Zhang, F. Liu, Y. Duan, A. Kangarlu and B.S. Peterson, "Improving the spectral resolution and spectral fitting of  $^1\text{H}$  MRSI data from human calf muscle by the SPREAD technique," *NMR Biomed*, vol. 27, p. 1325–1332, 2014.
- [16] Y. Zhang, Sh. Wang, P. Phillips, Z. Dong, G. Ji and J. Yang, "Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by

- PSOTVAC," *Biomedical Signal Processing and Control*, vol. 21, pp. 58-73, 2015.
- [17] Y. Zhang, Zh. Dong. Phillips, Sh. Wang, G. Ji and T. Yuan, "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," *Frontiers in Computational Neuroscience*, vol. 9, p. 66:1–15, 2015.
- [18] K. Maurer, S. Volk and H. Gerbald, "Augusted D and Alzeheimer's disease," *The Lancet*, vol. 349(9064), pp. 1546-1549, 1997.
- [19] A. Burns and S. Lliffe, "CLINICAL REVIEW Alzheimer's disease," *BMJ*, vol. 338, no. b158, pp. 467-471, 2009.
- [20] K. R. Gray, R. Wolz, R. a. Heckemann, P. Aljabar, A. Hammers, and D. Rueckert, "Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease," *Neuroimage*, vol. 60, no. 1, p. 221–229, 2012.
- [21] C. Cabral, P. M. Morgado, D. Campos Costa, and M. Silveira, "Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages," *Comput. Biol. Med.*, vol. 58, pp. 101-109, 2015.
- [22] H. Watanabe, M. Ono, and H. Saji, "Novel PET / SPECT Probes for Imaging of Tau in Alzheimer ' s Disease," *The Scientific World Journal*, vol. 2015, 2015.

- [23] H. Hanyu, T. Sato, K. Hirao, H. Kanetaka, T. Iwamoto, and K. Koizumi, "The progression of cognitive deterioration and regional cerebral blood flow patterns in Alzheimer's disease: A longitudinal SPECT study," *J. Neurol. Sci.*, vol. 290, no. 1–2, p. 96–101, 2010.
- [24] Y. J. Chen, G. Deutsch, R. Satya, H. G. Liu, and J. M. Mountz, "A semi-quantitative method for correlating brain disease groups with normal controls using SPECT: Alzheimer's disease versus vascular dementia," *Comput. Med. Imaging Graph.*, vol. 37, no. 1, p. 40–47, 2013.
- [25] F. Bloch, "Nuclear Induction," *Physical review*, vol. 70, no. 7-8, p. 460–474, 1946.
- [26] E.M Purcell,H.C Torrey and R.V Pound., "Resonance absorption by nuclear magnetic moments in a solid," *Physical review*, vol. 69, no. 1-2, pp. 37-38, 1946.
- [27] M. Symms, H.R Jager, K. Schmierer, T.A Yousry, "A review of structural magnetic resonance neuroimaging.," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, no. 9, pp. 1235-1244, 2004.
- [28] P. Jezzard,PM. Matthews,SM. Smith, "Functional MRI: an introduction to," Oxford University Press, Oxford , 2001.

- [29] Y. Zhang and S. Wang, "Detection of Alzheimer's disease by displacement field and machine learning," *PeerJ*, vol. e1251, no. 3, 2015.
- [30] E.S.A El-Dahshan, T. Hosny and A.B.M Salem, " Hybrid intelligent techniques for MRI brain images classification," *Digital Signal Processing*, vol. 20, pp. 433-441, 2010.
- [31] S. Chaplot, L.M Patnaik and N.R Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network.," *Biomedical Signal Processing and Control*, vol. 1, no. 1, pp. 86-92, 2006.
- [32] C. Plant, S.J Teipel, A. Oswald, C. Böhm, T. Meindl, J. Mourao-Miranda and M. Ewers, "Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease," *Neuroimage*, vol. 50, no. 1, pp. 162-174, 2010.
- [33] W. Zhu, N. Zeng and N. Wang, "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations," *NESUG proceedings, health care and life sciences, Baltimore, Maryland*, vol. 19, 2010.



- [34] M.H Zweig, & G. Campbell , "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *Clinical chemistry* , vol. 39, no. 4, pp. 561-577, 1993.
- [35] I. Dimitrovski, D. Kocev, I. Kitanovski, S. Loskovska, and S. Džeroski, "Improved medical image modality classification using a combination of visual and textual features," *Comput. Med. Imaging Graph*, vol. 39, pp. 14-26, 2015.
- [36] T. Xue, L. Bai, S. Chen, C. Zhong, Y. Feng, H. Wang, Z. Liu, Y. You, F. Cui, Y. Ren, J. Tian, and Y. Liu, "Neural specificity of acupuncture stimulation from support vector machine classification analysis," *Magn. Reson. Imaging*, vol. 29, no. 7, p. 943–95, 2011.
- [37] L.Y Hu,M.W Huang,S.W Ke and C.F Tsai., "The distance function effect on k-nearest neighbor classification for medical datasets," *SpringerPlus*, vol. 5, no. 1, p. 1304, 2016.
- [38] N. Speybroeck., " Classification and regression trees," *International journal of public health*, vol. 57, no. 1, pp. 243-246, 2012.
- [39] G. James, Majority vote classifiers: theory and applications, Stanford: Doctoral dissertation, Stanford University, 1998.

- [40] Q. Gao, J. Liu, H. Zhang, J. Hou, and X. Yang, "Enhanced fisher discriminant criterion for image recognition," *Pattern Recognit*, vol. 45, no. 10, p. 3717–3724, 2012.
- [41] H. Demirel and G. Anbarjafar, "Data fusion boosted face recognition based on probability distribution functions in different colour channels," *EURASIP J. Adv Signal Process*, vol. 2009, 2009.
- [42] I. Beheshti, "Structural MRI-based Classification of Alzheimer's Disease," Eastern Mediterranean University, Famagusta, 2016.
- [43] P. Bolourchi, M. Moradi, H. Demirel and S. Uysal , "Feature Fusion for Classification Enhancement of Ground Vehicle SAR Images.," in *UKSim-AMSS 19th International Conference on Modelling & Simulation*, 2017.
- [44] G. a. Papakostas, a. Savio, M. Graña, and V. G. Kaburlasos, "A lattice computing approach to Alzheimer's disease computer assisted diagnosis based on MRI data," *Neurocomputing*, vol. 150, pp. 37-42, 2015.
- [45] A. Savio, M. T. García-Sebastián, D. Chyzyk, C. Hernandez, M. Graña, A. Sistiaga, A. López de Munain, and J. Villanúa, "“Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI," *Comput. Biol. Med*, vol. 41, no. 8, pp. 600-610, 2011.

- [46] L. Khedher, J. Ramírez, J. M. Górriz, a. Brahim, and F. Segovia, "Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," *Neurocomputing*, vol. 151, pp. 139-150, 2015.