[UCI]I804:24011-200000265658

August 2016

Master's Degree Thesis

Alzheimer Disease Classification Using Lower Dimensional Features from Brain MRI and SVM Classifiers

Graduate School of Chosun University

Department of Information and Communication Engineering

Saruar Alam

Alzheimer Disease Classification Using Lower Dimensional Features from Brain MRI and SVM Classifiers

뇌 MRI 와 SVM 분류기로부터 낮은 차원의 특징을 이용한 알츠하이머 병 예측 기술

August 25, 2016

Graduate School of Chosun University

Department of Information and Communication Engineering

Saruar Alam

Alzheimer Disease Classification Using Lower Dimensional Features from Brain MRI and SVM Classifiers

Advisor: Prof. Goo-Rak Kwon

This thesis is submitted to Chosun University in partial fulfillment of requirements for a master's degree

April 2016

Graduate School of Chosun University

Department of Information and Communication Engineering

Saruar Alam

알람 사루아르의 석사학위논문을 인준함

위원장	조선대학교	교수	변재영	(인)
위 원	조선대학교	교수	김영식	(인)
위 원	조선대학교	교수	권구락	(인)

2016년 5월

조선대학교 대학원

Contents

List of Tables

List of Figures

Acronyms

AD	**Alzheimer Disease**
ADNI	**Alzheimer's Disease Neuroimaging Initiative**
CAD	**Computer Aided Diagnosis**
CDR	**Clinical Dementia Rating**
CWT	**Complex Wavelet Transform**
DWT	**Discrete Wavelet Transform**
DtCWT	**Dual Tree Complex Wavelet Transform**
HC	**Healthy Controls**
LDA	**Linear Discriminant Analysis**
MCI	**Mild Cognitive Impairment**
Mkl-SVM	**Multi Kernel Support Vector Machine**
KPCA	**Kernel Principal Component Analysis**
MRI	**Magnetic Resonance Image**
MRMR	**Maximum Relevance Minimum Redundancy**
NC/CN	**Normal Controls**
PCA	**Principal Component Analysis**
SVM	**Support Vector Machine**
WT	**Wavelet Transform**

Abstract

Alzheimer disease prediction using lower dimensional features from brain MRI and SVM classifiers

Saruar Alam
Advisor: Prof. Goo-Rak Kwon
Associate Professor
Dept. of Information and
Communication Engineering
Chosun University

Alzheimer's disease (AD) is the most frequent form of dementia, causes more health and socioeconomic problem. This progressive neurodegenerative disorder causes damage of brain cells, affects cognitive acts, behavioral problems, and memory disorder. At present, a lot of computer-aided diagnosis (CAD) based research is going on using Magnetic Resonance imaging (MRI) as a biomarker. Early diagnosis could help the patient to take preventive cure to get rid of AD risk factors to generate further. Around 50% MCI (Mild Cognitive Impairment) patient develop AD in three to four years. In this paper, a novel method is proposed to predict AD from normal controls (NC) here. In this work, maximum relevant minimum redundant principal components of dual tree complex wavelet transform (dtCWT) coefficients are extracted. The transaxial slices of MR images are selected for extracting dtCWT coefficients here. After linear discriminant analysis (LDA) of those coefficients, kernel SVM is trained and tested. The accuracy, sensitivity, and specificity we have achieved using proposed approach are comparable or superior to those obtained by various conventional AD prediction methods found in the literature.

초록

뇌 MRI 와 SVM 분류기로부터 낮은 차원의 특징을 이용한 알츠하이머 병 예측 기술

알람 사루아르

지도교수 : 권구락 교수

정보통신공학과

조선 대학교 대학원

알츠하이머성 치매는(AD) 치매의 가장 흔한 형태이며, 치료 비용으로 인한 경제적 문제와 더불어 사화적 문제가 발생한다. 이는 뇌 세포 손상으로 인한 신경 퇴행성 장애로서 인지장애, 행동장애, 기억장애를 유발한다. 현재 컴퓨터를 이용한 진단으로써 많은 연구가 이루어지며 대표적으로 자기 공명 영상(MRI)의 바이오마커를 이용하여 진단하는 방법이 있다. 치매 조기 진단은 알츠하이머성 치매의 초기치료를 통해 추가적으로 환자가 늘어나지 않게 조치 할 수 있다. 50%의 MCI(경도인지장애) 환자는 3~4 년 동안 알츠하이머성 치매가 진행된다. 본 논문에서는 정상인 대조군(NC)에서 알츠하이머성 치매를 예측하는 방법을 제안한다. 본 연구에서는, dtCWT(듀얼 트리 콤플렉스 웨이블릿 변환)의 최대에서 최소한의 중복 주성분 계수를 추출한다. 뇌 MRI 영상의 시상면, 관상면, 가로면(Transverse Plane) 영상은 dtCWT 계수를 추출하기 위해 사용된다. 이후 계수들은 LDA(선형 판별 분석)를 거쳐 Kernal-SVM 학습을 통해 트레이닝되고 테스트된다. 이 논문에서 제안하는 방법은 기존의 다양한 연구들에서 찾을 수 있는 AD 예측방법과 정확성, 민감도, 특이도를 따졌을 때 더욱 우수하거나 대등한 성능을 보인다.

Chapter 1

Introduction

1.1 Thesis motivation and overview

Alzheimer disease (AD) is the most frequent form of dementia, for which neurological disorder, declination of mental state occurs. Its wide range of symptoms affects the patient to carry out normal daily activities. The disease causes loss of function, metabolic alterations and structural changes in the brain. It is not just an ordinary disease, but significantly 6th leading cause of death in USA, for which no significant cure or effective treatment is currently manifested. AD is major cause of dementia, a growing health and socioeconomic problem, due to the progressive ageing of the world population. Early diagnosis of Alzheimer disease (AD) and Mild Cognitive Impairment (MCI) is always constitutive. The preventive measure might get an impact to degenerate AD risk factors. Earlier diagnosis method was based on clinical observation and cognitive evaluation. Recently, Many Computer-Aided Diagnosis (CAD) based methods have been studied to predict AD/MCI from Healthy Controls. Structural Magnetic Resonance Imaging (sMRI), one of the most important biomarker widely used to visualize the changes in brain morphology as it is non-invasive. Using MR images, it can be distinguished among different brain tissues, such as Gray Matter (GM), White Matter (WM), and CerebroSpinal Fluid (CSF). One of most popular research initiatives for early detection and tracking AD progression is the Alzheimer ’s disease Neuroimaging Initiative (ADNI) which is an ongoing, longitudinal, multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers.

Due to the widening interest in this technique, lots of researchers around the world are keen for CAD based AD/MCI prediction, and several research organizations have started neuroimaging recruiting. The discovery of robust CAD based stratification of AD/MCI and robust image biomarkers will let the researchers to focus treatments on the early affected regions. It will also assist to make a diagnosis before the disease symptomatology appears.

1.2 Goals of this work

In this context, the main aim is to propose accurate noble method to classify AD from HC using brain MR image as biomarker. Our goal is to use dual tree wavelet based feature from extracted 2D MR image slices, and also use segmented volumetric feature from 3D MRI. ONIS Viewer is used to extract 2D MR image slice from 3D MRI. To extract segmented volumetric feature, FreeSurfer tool is used. For The details about preprocessing and feature extraction are explained in section 2.3. The kernel SVM will be fed with the extracted feature after reducing dimension and analyzing the discrimination.

The 2D MR image slices are extracted and subjectively selected. To extract almost shift invariant wavelet based coefficients, dual tree complex wavelet transform (dtCWT) is applied on those 2D MR image slices. After discarding less relevant and high redundant dtCWT coefficients, and reducing the dimension, kernel SVM is be trained and tested.

1.3 Contribution

According to best of my knowledge, a noble approach is applied in this work. In this method, subcortical segmented volumetric feature is extracted. The dual tree complex wavelet transform is applied on selected 2D MR image slices to extract almost shift invariant feature, maximum relevance and minimum redundant feature subset is selected, and principal component analysis is used for dimensionality reduction. The principal components are projected into another dimensional space applying Fisher Linear Discriminant Analysis (Fisher LDA) to ensure variability of features within and between classes, and then kernel SVM is trained and tested on linear discriminant of principal components of max-relevant min redundant dtCWT coefficients.

A simulation code has been programmed in Matlab 2015b to analyze the prediction performance with reference to accuracy, sensitivity, specificity, and compared with other conventional AD classification methods

1.4 Thesis Structure

The remaining chapters of the paper is arranged as follows: in chapter 2, it will reviewed about Alzheimer Disease, MR Image processing and feature extraction, briefly about feature subset and dimensionality reduction technique, and mathematical formulation of linear Support Vector machine classifier.

Then, mathematical formulation lying behind dtCWT based coefficients extraction, feature subset selection, dimensionality reduction, classification problem of our proposed method will be summarized in Chapter 3. Then, the

experimental results will be analyzed in Chapter 5. Finally the conclusion and the future research possibilities will be presented in Chapter 6.

Chapter 2

Background

2.1 Alzheimer's disease

The Alzheimer disease (AD) is a most frequent type of dementia, primarily suffered by the elders. The condition is named for Alois Alzheimer, a German psychiatrist who was the first credited with identifying and describing the conditions, later linking its symptoms and pathology in 1906 [1]. A progressive neurodegenerative disorder, AD, instigates to damage brain cells, and then induces cognitive assessment, behavioral dilemma, and memory disarray. It is a progressive neurological disorder, gradually it worsens over time, for which there is currently no cure, leading eventually to death, however promising research and development for early detection and treatment is underway. The percentage of people over 84 years old suffering from the disease is up to 42% [2]. According to statistical report, it is predicted that above 35 million people worldwide will suffer from dementia by 2030, is triple than the number of patients who develop at present. The ubiquity of AD varies among many different factors, including age, co-morbidities, genetics, and education level. The diagnosis cost of AD patients is estimated to $220 billion per year and $ 605 billion per year globally. The cerebral atrophy in entorhinal cortex and hippocampus is instigated by senile plaque which consists of amyloidal beta-42 protein. Neurofibrillary tangle consists of tau protein also causes atrophy at hippocampus. Early diagnosis of AD or MCI is always indispensable because it can possibly help to control the development of AD.

AD progresses over time. There are mainly three stages of the disease progression. At each stage the symptoms and challenges are different. Unique set of symptoms varying in severity could be observed at each stage of AD. While identifying current stage, the physician can predict its next stage and follow up symptoms, and purveys possible course of treatment accordingly.

Mild Alzheimer's disease (early-stage):

The early stage lasts 2 to 4 years. In this stage, gradual decline of patient's cognitive ability can be realized. Common symptoms at this stage include [3]:

- Problems presenting the right word or name
- Trouble memorizing names when introduced to new people
- Having greater difficulty performing tasks in social or work settings
- Failing to remember material that one has just read
- Losing or misplacing a useful object
- Increasing trouble with planning or arranging

Moderate Alzheimer's Disease (Middle Age):

This stage endures for 2 years, longest stage, more difficulties are experienced by AD patients. The patient needs special assistance of others to carry on daily living activities. The most common symptoms reported at this stage include [3]:

- Forgetfulness of incident or about one's own personal history
- Feeling moody or withdrawn, especially in socially or mentally challenging circumstances

- Being unable to recall their own address or telephone number or the high school or college from where they graduated
- Confounder about where they are or what day it is
- The need for help choosing proper clothing for the season or the occasion
- Trouble controlling bladder and bowels in some individuals
- Alteration in sleep patterns, such as sleeping during the day and becoming restless at night
- An increased danger of wandering and becoming lost
- Personality and behavioral changes, including suspiciousness and deception or compulsive, repetitive behavior like hand-wringing or tissue shredding

Severe Alzheimer's disease:

The severe stage lasts between 1 to 3 years. In this stage, cognitive functioning continues to decline and physical condition gets worse. At this stage, individuals may:

- Require full-time, around-the-clock assistance with daily personal care
- Lose awareness of recent experiences as well as of their surroundings
- Require high levels of assistance with daily activities and personal care
- Experience decline in physical abilities, including the ability to walk, sit, and eventually swallow
- Have increasing difficulty in communication
- Become vulnerable to infections, especially pneumonia

The Clinical Dementia Rating (CDR) is a numeric scale, can be described to quantify dementia rating based on different set of severity. The patient's cognitive and functional performance is graded numerically based on six areas: memory, orientation, judgment & problem solving, community affairs, home & hobbies, and personal care [4]. The numerical scores of each area are integrated to get composite score of each patient ranging from 0 through 3.

The qualitative equivalence of CDR scaling are shown in Table 1.1.

Composite Rating	Symptoms
0	none
0.5	very mild
1	mild
2	moderate
3	severe

Table 1.1: CDR of different stage of severity.

2.2 Magnetic Resonance Imaging

The biomarker, structural Magnetic Resonance Imaging (sMRI) can be used for brain atrophy measurement to identify abnormal volumetric changes related to AD [22]. MRI employs the phenomenon of nuclear magnetic resonance (NMR) to produce high quality structural images of the internal organs and other tissues. Magnetic resonance imaging (MRI) technique was first used in 1977. It creates two or three dimensional images of the body that can be used to diagnose abnormalities, injury, and illness. The vital component of the MRI system is the superconducting magnet, which

produces a large and stable magnetic field [5]. The smaller gradient magnets are there to create weaker magnetic fields. These magnets allow for distinct parts of the entire body to be scanned. The human body constitutes of billions of atoms. However, it is the hydrogen atoms that are responded by the magnetic field. Each hydrogen atoms are randomly spinning around an axis, but inside the magnetic field of the MRI, the molecules are lined up with the orientation of the field. Half of the atoms direct towards the patient's head, and rest half direct towards the feet, cancelling each other out. A few atoms out of every million are left out without cancelling out. Then the machine emits a radio frequency (RF) pulse specific to hydrogen, which causes these protons to spin towards a different point. When the spinning breaks off, the protons release energy, which is defined by the system. Using a contrast dye, each type of tissue responds differently and appears as a unique intensity of gray color when the image is created [6].

Knowing how the system works, researchers are able to determine if an MRI can effectively ascertain the occurrence of the structural changes and cellular death seen in the brain of an AD patient. Atrophy of the hippocampus is often viewed in AD, even before the appearance of clinical symptoms [7]. The Nun Study, conducted in 2002, collected postmortem MRI scans of 56 participants with varying stages of cognitive impairment. The MRI was used to detect the hippocampal volume and determine its significance as a biomarker of AD neuropathology [8]. Temporal lobe atrophy is closely associated with the development of AD [9], and histological studies show that the hippocampus, amygdala and entorhinal cortex are particularly sensitive to AD pathology [10]. Correlation has been found between the rate of temporal lobe atrophy and both current cognitive performance and future cognitive decline, even among normal healthy individuals [11]. Increased

rates of hippocampal atrophy compared with cognitively healthy individuals have been measured using MR image in both AD and MCI patients [12, 13]. Longitudinal studies have additionally shown that the rate of hippocampal atrophy beefs up over time in both AD and MCI patients [14, 15]. However, hippocampal atrophy alone is not sufficient to classify conversion from MCI to AD, and other structures may prove more sensitive [16].

Several ADNI MRI data studies found that an increased rate of hippocampal volume loss was associated with presence of the ApoE e4 allele in AD patients, and with reduced levels of CSF Aβ in MCI patients [12]. Another study showed that the rate of temporal lobe atrophy in AD is associated with reduced CSF Aβ and elevated CSF tau, and that it is significantly faster in MCI subjects that later develop to AD than in non-converters [17]. The result and discussion prompted that the scans could be used to identify non-demented elderly with AD neuropathology who have not yet presented with memory impairment. By identifying the risk for these patients to convert AD well before the appearance of symptoms, physicians may be able to carry out treatment to slow the progression of the disease.

A recent study conducted in 2009 by the Departments of Radiology and Neurology at the University of Pennsylvania explored the use of sodium magnetic resonance imaging in the development of AD. This imaging technique uses the same principle as illustrated above. However, instead of measuring the hydrogen atoms, naturally abundant sodium, Na was used by this technique [18]. This ion was chosen for this because of the capability of sodium in the brain to detect tumors and track cell death [19]. The participants included five healthy normal elderly adults and five who had a probable diagnosis of AD. When neuronal death occurs, the intracellular

space is shrunk. Therefore, there is an increased concentration of sodium in the extracellular space, generating stronger signal intensity from the MRI for patients who have AD. Though this technique is not yet perfected, studies are being conducted to determine if the increased signal intensity is induced by a change in ion concentration or a change in volume [20].

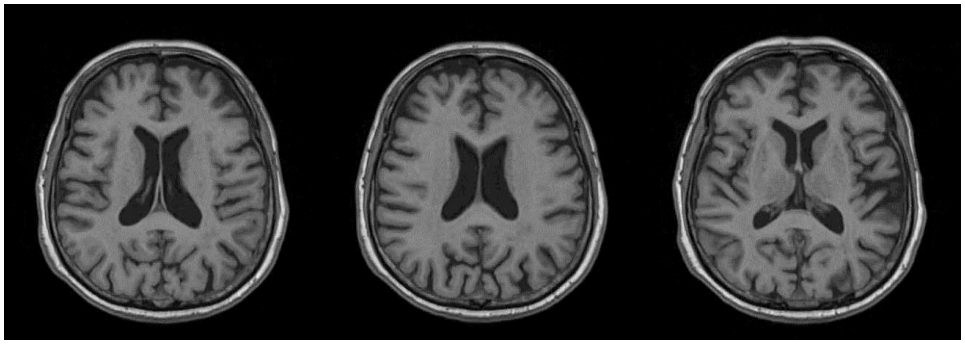
MRI Pros and Cons:

When taking the effectiveness of this technique into consideration, there are both pros and cons. Potential advantages of choosing this procedure are that it is painless, noninvasive, and can detect very minute abnormalities without the radiation exposure of an X-ray. The resulting reconstructed image also has high spatial resolution. However, this process is very expensive and may not be covered by insurance. The space inside the scanning machine is very small, which may make it hard to examine a claustrophobic patient. If a patient has metallic objects inside of their body, they cannot use the MRI system due to high sensitivity to strong magnetic field.

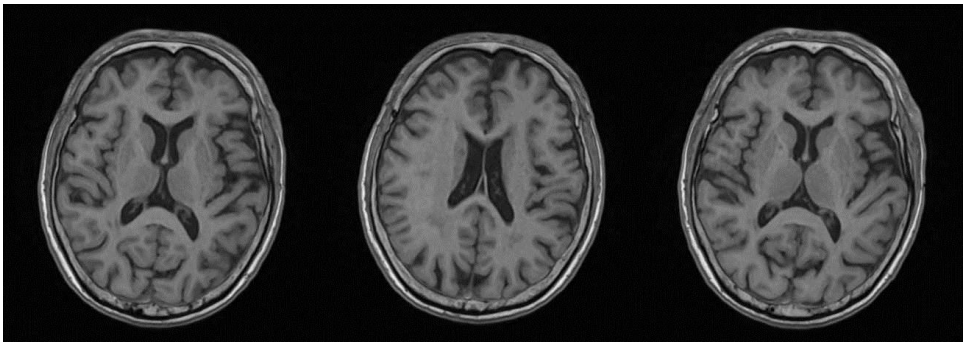
MRI Accuracy:

A study conducted by the Florida Alzheimer's Disease Research Center found that MR image scans are effective in detecting the brain atrophy observed in AD. They collected brain MRI for 260 participants, some with mild cognitive impairment, others with probable AD, and a control group of elderly healthy adults with no memory decline. The researchers effectively matched the scans with the correct group of patients based on the amount of atrophy in the mid-brain. Some scans showed brain atrophy before any symptoms were present, hinting that this technique would be effective for

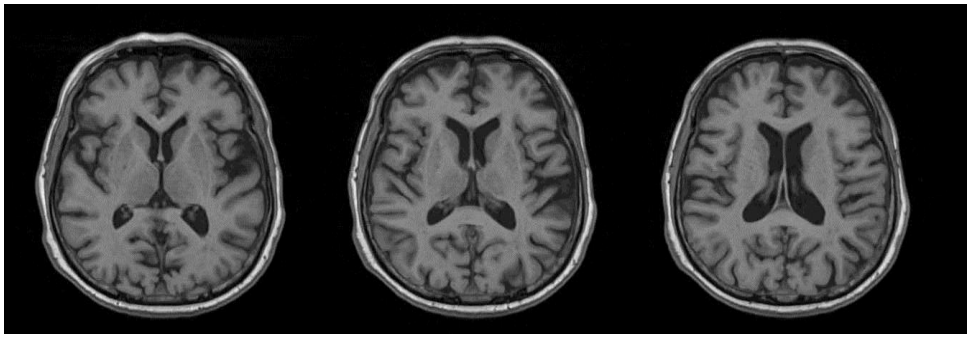
early diagnosis of the disease [21]. Sample transverse slices from MR images of healthy individuals, Mild Cognitive Impairment, and AD patients are depicted in Fig 2.1. These images hint that AD patients typically show evidence of cortical atrophy, and enlarged ventricles in comparison with healthy controls.



(a)



(b)



(c)

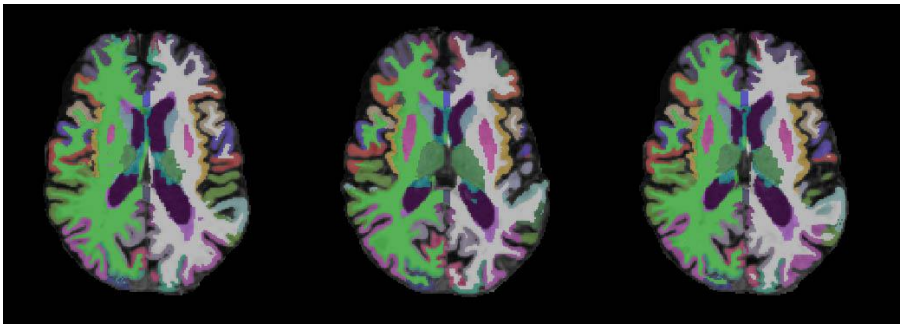
Fig. 2.1: Transverse slices from MR images of (a) healthy individuals, (b) Mild Cognitive Impairment, and (c) AD patients. These images indicate that AD patients typically show evidence of cortical atrophy, and enlarged ventricles in comparison with healthy controls.

2.3 MR Data Preprocessing

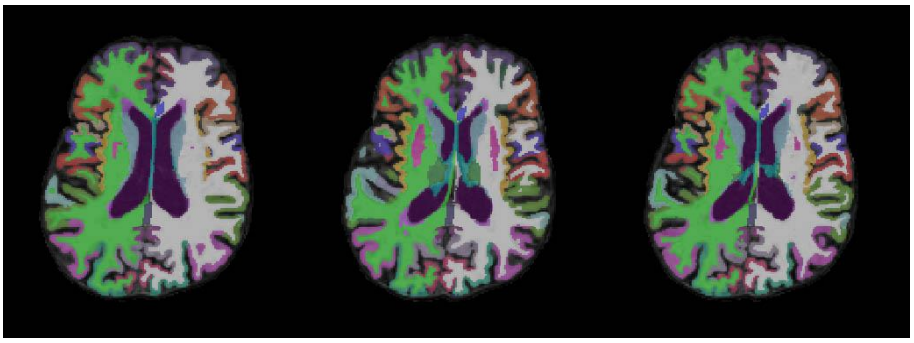
2.3.1. FreeSurfer

Freesurfer [23] is most broadly used software for analysis of structural and functional imaging biomarkers. It is fully automatic brain imaging tool pipeline used for displaying of cortical surface between white and gray matter, representation of the pial surface, segmentation of white matter from the rest of the brain, skull stripping, B1 bias field correction, nonlinear registration of the cortical surface of an individual with a stereotaxic atlas, labeling of regions of the cortical surface, statistical analysis of group morphometric differences, and labeling of subcortical brain structures etc. Fischl [23] demonstrated the implementation of complex image processing pipeline and the subsequent computation of corresponding volume of segmented plenty numbers of image scans of anatomical structures. Freesurfer has the disadvantage of taking more time for processing as compared to Statistical Parametric Mapping (SPM) tool. By using Freesurfer total GM, the TIV (called intracranial volume), and hippocampus and ventricular volumes were obtained directly from the asegstats output file, left and right temporal GM volumes were extracted by summing up several ROI volumes found in the lh.aparc.stats and rh.aparc.stats files, respectively. 34 gyral based cortical ROIs are detected [24] at each hemisphere from human cerebral cortex. In automatic subcortical segmentation, each voxel in the normalized brain volume is assigned one of about 40 labels [25], including: Cerebral White Matter, Cerebral Cortex, Lateral Ventricle, Inferior Lateral Ventricle, Cerebellum White Matter, Cerebellum Cortex, Thalamus, Caudate, Putamen, Pallidum, Hippocampus, Amygdala, Lesion, Accumbens area, Vessel, Third Ventricle, Fourth Ventricle, Brain Stem, Cerebrospinal Fluid.

It also registers the individual cortex surfaces to surface-based anatomical atlases (Desikan-Killiany, Destrieux) [26, 27]. Techniques are implemented for automatically assigning a neuroanatomical label to each region on a cortical surface model based on probabilistic information estimated from a manually labeled training set (implemented using FreeSurfer). This procedure incorporates both geometric information derived from the cortical model, and neuroanatomical convention, as found in the training set. The result is a complete labeling of cortical sulci and gyri. A subcortically segmented 2D MRI slice of AD and healthy patient have been shown in Fig. 2.2.



(a)



(b)

Fig. 2.2: Subcortically segmented Transverse slices from MR images of (a) AD patients, and (b) healthy individuals by using FreeSurfer tool.

2.3.2. Discrete Wavelet Transform

The discrete wavelet transform (DWT) is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations obeying some defined rules. The DWT captures both frequency and location information (location in time). It is an effective mathematical tool for feature extraction, its wavelet coefficients can be used as feature of MR image. The scale parameters are discretized for practical reason. The scale parameter, d is quantized with respect to translation parameter, δ as shown below,

$$d = 2^{-m} \quad and \quad \delta = n2^{m},\ where\, m, n \in Z\,. \qquad (2.1)$$

Thus, the family of wavelet functions can be denoted as

$$\psi_{m,n}(t) = 2^{m/2}\psi(2^{m}t - n)\,. \qquad (2.2)$$

The DWT segregates input signal $\mathrm{x}(t)$ into a set of synthesis wavelets as depicted in Eqs. (2.3) and (2.4),

$$\mathrm{x}(t) = \sum_{m}\sum_{n} q_{m,n}\psi_{m,n} \quad , \qquad (2.3)$$

where, $q_{m,n} = \langle x(\mathrm{t}), \psi_{m,n}(\mathrm{t})\rangle$.

Considering discrete signal x[n], the wavelet segregation on J octaves is shown by

$$x[\mathrm{n}] = \sum_{i=1 to J}\sum_{k\in Z} q_{\mathrm{i,k}} g[\mathrm{n} - 2^{i}\,\mathrm{k}] + \sum_{k\in Z} u_{J,\mathrm{k}} h_{J}[\mathrm{n} - 2^{J}\,\mathrm{k}]\,, \qquad (2.4)$$

where, the wavelet coefficients $q_{\mathrm{i,k}}, i = 1, \ldots, J$, and scaling coefficients $u_{\mathrm{J,k}}, i = 1, \ldots, J$.

$$q_{\mathrm{i,k}} = \sum_n x[\mathrm{n}]\mathrm{g}_i^*[\mathrm{n} - 2^i\,\mathrm{k}], \tag{2.5}$$

$$d_{\mathrm{i,k}} = \sum_n x[\mathrm{n}]\mathrm{h}_J^*[\mathrm{n} - 2^J\,\mathrm{k}],$$

where, $\mathrm{g}_i[\mathrm{n} - 2^i\,\mathrm{k}]$ and $\mathrm{h}_J[\mathrm{n} - 2^J\,\mathrm{k}]$ denotes discrete wavelets and scaling sequences respectively, (*) indicates complex conjugate.

The DWT can be represented in 2D image also. The resultant of image S is breaking into first level of low approximation coefficient S_a^1 , detailed horizontal component S_h^1, and detailed vertical component S_v^1, S_d^1 [28]. The approximation coefficient S_a consist of low frequency components, and S_v^1 and S_d^1 contain high frequency components.

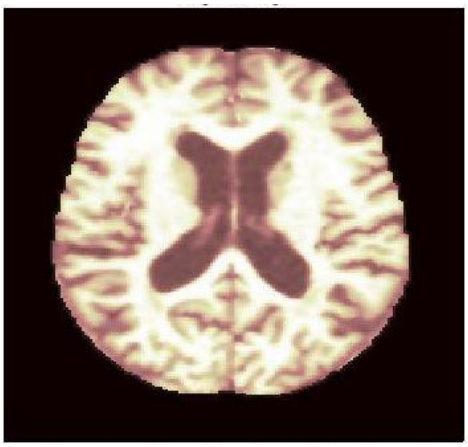
Thus image can be represented as follows:

$$S = S_a^1 + S_h^1 + S_v^1 + S_d^1. \tag{2.6}$$

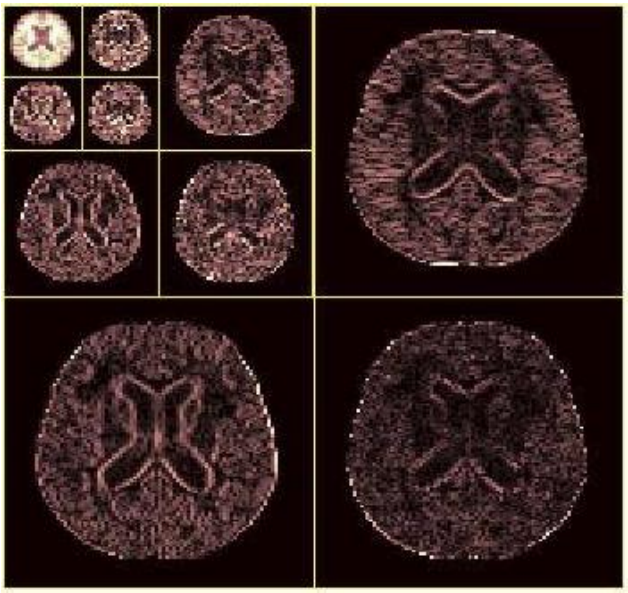
Similarly image can be decomposed into multi scale wavelets. The particular level approximation and detailed components are obtained applying DWT in previous level approximation component. If the process is repeated up to L level, the image S can be represented as $L\,th$ terms approximation component (S_d^L), and detailed components as depicted below in Eq. (2.7),

$$S = Y_a^N + \sum_{i=1\,to\,L} \{S_h^i + S_v^i + S_d^i \quad . \tag{2.7}$$

The dimension of decomposed image is reduced by half at each decomposition level. Hence the dimension of all detailed components achieved from first level of decomposition of an $N*N$ image is $N/2*N/2$, second level is $N/4*N/4$, and so on. The more the level increases, the more the image gets compact as shown in Fig.2.3. Thus a simple hierarchical framework is obtained for interpreting image transform [29].



(a)



(b)

Fig. 2.3: Transverse slices from MR images of (a) Original Image for wavelet transform, and (b) Detailed components of 1st, 2nd, 3rd level decomposition.

2.3.3. Feature selection

In this work, 2 sample t test is used to discard features, hence it reduces dimension also. It calculates a confidence interval and executes a hypothesis test of the difference between two population means when sample distribution is independent to each other and standard deviations are unknown. Confidence level in the result increases when the sample size increases. If sample size is small, it works best if the data in normally or close to normally distributed.

The two-sample t-test compares the location parameter of two independent data samples.

The test statistic is

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}, \tag{2.8}$$

where, $\bar{x}$ and $\bar{y}$ are the means of samples, n and m are the sample sizes, and s_y and s_x are sample standard deviations.

If it can be assumed that the two samples having equal variance from populations, the test statistic under the null hypothesis has Student's t distribution with $n+m-2$ degrees of freedom, and then the sample standard deviations are substituted with the pooled standard deviation as shown below.

$$s = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}, \tag{2.9}$$

The significance level of null hypothesis can be determined by using two sample t test.

2.3.4. Dimensionality reduction technique

Higher dimensionality of feature is a curse for pattern classification. For smooth classification, dimensionality reduction technique transforms data from higher dimensional space to lower dimensional space. The data transformation may be linear or nonlinear. One of most frequently used linear transformation is Principal Component Analysis (PCA). PCA is orthogonal transformation to convert possibly correlated samples into linearly uncorrelated variables. The number of principal components is less than or equal to the number of original variables. The PCA is summarized as

- Finding mean of the data matrix and zero mean matrix.
- Constructing covariance matrix.
- Getting the eigenvalue and the eigenvector.
- Projecting the data matrix with eigenvectors corresponding to highest to lowest eigenvalues.

2.4 Support Vector Machine

Support Vector Machine, basically a binary classifier is very efficient for classification of both linearly separable and non-separable data. It finds best hyper plane that separates both the classes having optimum margin from support vectors during training phase as portrayed. While testing with new data point, classifier takes decision on the basis of hyper plane.

Let consider the training set $\{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in R^d$ training sample is and its corresponding class level is $y_i \in \{-1, +1\}$. It is needed to find maximum-margin hyperplane that divides one group (+1) from other (-1).

The hyperplane can be represented as

$$\vec{w}.\vec{x} - b = 0, \tag{2.10}$$

where, $\vec{w}$ is weight vector which is normal to hyperplane, b is bias term.

Hard-margin:

If the data samples are linearly separable, then two parallel hyperplane can be drawn to classify two classes of data so that distance between them is optimum. The region bounded by these two hyperplanes is called the "margin".

The optimization problem can be represented as

$$\begin{aligned} &\text{minimize } \|\vec{w}\| \\ &\text{Subject to } \mathrm{y}_i(\vec{w}.\vec{x} - b = 0) \geq 1, \ \ for\ i = 1, \ldots, \mathrm{n}, \\ &n \in Z \end{aligned} \tag{2.11}$$

The decision function is

$$\mathrm{f(x)} = \mathrm{sign}(\vec{w}.\vec{x} - b = 0). \tag{2.12}$$

Soft-margin:

If the samples are not linearly separable, SVM optimization problem is extended introduction hinge loss function without using kernel trick as

$$\left[\frac{1}{n}\sum_{i=1}^{n}\max(0,1-\mathrm{y}_i(\vec{w}.\vec{x}-b))\right]+C\,\|\overrightarrow{w}\|, \qquad (2.13)$$

where, $\max(0,1-\mathrm{y}_i(\vec{w}.\vec{x}-b))$ is hinge loss function, C is trade of parameter to get optimum margin to ensure $\vec{x}_i$ to be on the right side of class or hyperplane. If there are small value of trade of parameter C and linearly separable data samples, soft-margin SVM will behave indistinguishably like hard-margin SVM.

Chapter 3

The Proposed Method

3.1 Overview

Early preventive care could help to degenerate its risk factors gradually. The noninvasive biomarker Magnetic Resonance (MR) images are used here because morphometric difference and cerebral atrophy could be realized. MR image slices are used after to extract feature. A novel approach is applied for predicting AD from HC using complex dual tree wavelet coefficients extraction, principal components of min-redundancy and max relevance feature subset selection, and linear discriminant analysis. The method is compared with segmented volumetric feature based AD prediction where FreeSurfer is used for feature extraction. By using Freesurfer, the TIV (called intracranial volume), total GM, and hippocampus and ventricular volumes were obtained directly from the asegstats output file, left and right temporal GM volumes were extracted by summing up several ROI volumes found in the lh.aparc.stats and rh.aparc.stats files, respectively. 34 gyral based cortical ROIs are identified [24] at each hemisphere from human cerebral cortex.

3.2 Dual tree complex wavelet transforms

Wavelet is one of most frequently used feature extraction technique for MR images. The advantage of WT over Fourier transform is its multiple scaled representations, frequency components with spatial domain information. Fourier coefficients only purvey frequency information of

image, but Wavelets contains powerful observation of both spatial and frequency domain in multiple scaled format. Wavelets representation is spatially localized in space, but Fourier functions are not spatially localized because it consists of only frequency components of image. MR images can be represented and processed at various resolutions; hence it can be used as incisive framework for processing multi resolution images. The DWT coefficients can be extracted by using array of low and high pass filter banks.

But conventional wavelet transformation has multiple drawbacks. The disadvantage includes drift in wavelet coefficient oscillation towards positive and negative around singularities, shift variance of signal which may cause oscillation of wavelet coefficient sample around singularities, the substantial aliasing of amply spaced wavelet coefficients patterns, and lack of directional selectivity perturbs to process and model geometric image features like edges and ridges. Then flaws regarding conventional DWT are not experienced by Fourier transform. To overcome the drawbacks, the improved dual tree complex wavelet transform (dtCWT) is used [30]. Getting inspired by Fourier transform, CWT can be represented as complex-valued scaling function and complex-valued wavelets. Dt-CWT engages two real DWTs; the first DWT purveys the real part of the transform and second SWT provides the imaginary part. Two kinds of filter set, analysis filter banks and synthesis filter banks are used for implementing dual tree wavelet transform to safeguard that overall transformation becomes approximately analytic as show in Fig. 3.1.

The dt-CWT can be denoted in matrix form

$$D = [\mathrm{D}_h \ \mathrm{D}_g], \tag{3.1}$$

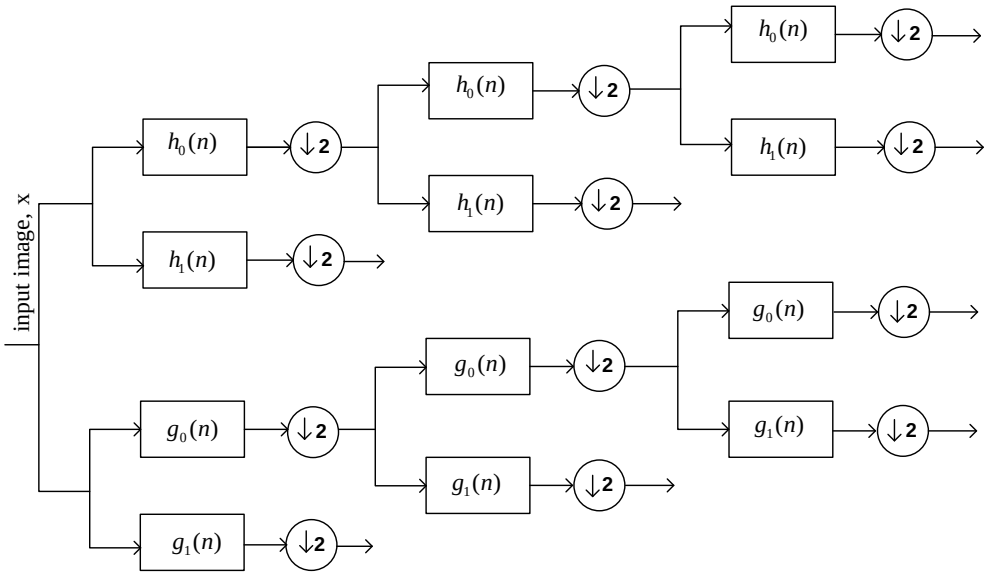


Fig. 3.1: Block diagram for a 3-level DTCWT.

where, D_h, D_g are rectangular matrices.

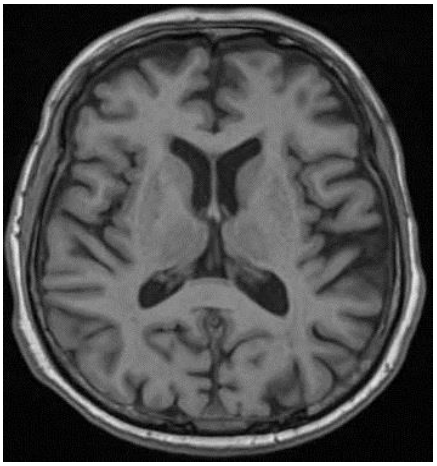
The input image x, its complex wavelet coefficients can be represented as

$$T_h + jT_g \tag{3.2}$$

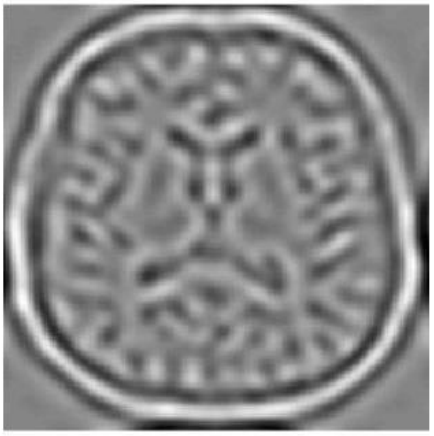
where, $T_h = D_h * x$ is the real part, and $T_g = D_g * x$ is the imaginary part.

The dt-CWT coefficients of input image are shift invariant because it does not change when the image is shifted in time and space domain. Dt-CWT employs segregation of 6 different directions (±15, ±30, ±45) for 2D image, 28 different direction for 3D image while conventional DWT only allows isolation of horizontal and vertical direction . For each 2D slice of a subject, 4-level dtCWT coefficients are extracted in one particular scale. In Fig. 3.2, it has been shown the reconstructed image, and error after DWT and DtCWT coefficients of a 2D sample slice. When compared both the reconstructed

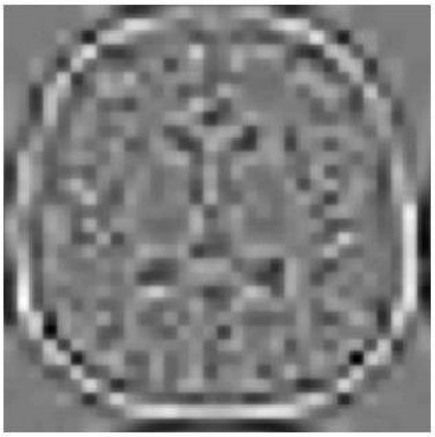
images, the dual tree complex wavelet transform more faithfully reproduces line and curve singularities.



(a)



(b)



(c)

Fig. 3.2: Transverse slices from MR images of (a) Original image for reconstruction, and (b) Reconstructed image from dtCWT, reconstructed error 0.7326 (c) Reconstructed image from DWT, reconstructed error 0.8812.

3.3 Max-Relevance and Min-Redundancy feature subset

Higher dimensional feature is a curse for pattern classification. The entire feature doesn't purvey disease signature. The dimension is reduced selecting feature subset based on maximum dependency, max-relevance, and min redundancy [31].

Let $\{x_i\}$ is n dimensional feature, S is subset of m features of x_i having the maximum dependency towards target class or level l. The max-dependency can be expressed as

$$\max D(S,l), \quad D = I(\{\mathrm{x}_i; \mathrm{i} = 1,...,\mathrm{m}\}; \mathrm{c}) . \tag{3.3}$$

The features are added one by one, and last feature is selected which have highest contribution to the increase of $I(\mathrm{S}_m;\mathrm{l})$ as shown in Eq. (3.4).

$$\begin{aligned} I(\mathrm{S}_m;\mathrm{l}) &= \iint p(\mathrm{S}_{m,}\mathrm{l}) \log \frac{p(\mathrm{S}_m,\mathrm{l})}{p(\mathrm{S}_m)\,p(\mathrm{l})} dS_m, \\ &= \iint p(\mathrm{S}_{m-1,}\, x_m,\mathrm{l}) \log \frac{p(\mathrm{S}_{m-1},\mathrm{x}_m,\mathrm{l})}{p(\mathrm{S}_{m-1},\mathrm{x}_m)\,p(\mathrm{l})} dS_{m-1} dx_m dl, \\ &= \int ....\int p(x_1,...,x_m,\mathrm{l}) \log \frac{p(x_1,...,x_m,\mathrm{l})}{p(x_1,...,\mathrm{x}_m)\,p(\mathrm{l})} dx_1...dx_m dl, \end{aligned} \tag{3.4}$$

As max-dependency has a drawback of slow computational speed, maximum relevance features are chosen with mean value of all mutual information between x_1 and its assigned level l.

$$\max\ \mathrm{D(S,l)},\ \mathrm{D} = \frac{1}{|\mathrm{S}|} \sum_{x_i \in S} I(x_i;l) . \tag{3.5}$$

The extracted relevant feature may have high redundancy. To discard similar kind of features, min Redundancy is formulated as

$$\min\ R(\mathrm{S}),\ R = \frac{1}{|\mathrm{S}|^2} \sum_{x_i, x_j \in S} I(x_i; \mathrm{x}_j). \tag{3.6}$$

The norm of incorporating above two constraints is called min-redundancy and max-relevance (mRMR), $\psi(D, R)$. The simple optimization of D and R simultaneously can be defined as

$$\max\ \psi(D, R),\ \psi = D - R\ . \tag{3.7}$$

Selection of S_m -*th* from nearest S_{m-1} -*th* feature $\{\mathrm{X} - \mathrm{S}_{m-1}\}$ is set by incremental search algorithm maximizing $\psi(.)$. The respective incremental search method is optimized as follows:

$$\max_{x_j \in X - S_{m-1}} \left[I(x_j; l) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_j; x_i) \right]. \tag{3.8}$$

3.4 Principal component Analysis

It is a dimensionality reduction technique [32] as higher dimensional feature has the disadvantage for pattern classification. Extracting principal components are illustrated mathematically below,

Let $X\, be\, N*d$, (N=number of patients, d-dimensionality) dimensional imaging feature,

After getting data into zero centered,

$$\frac{1}{N}\sum_{i=1}^{N} x_i = 0 \quad . \tag{3.9}$$

Diagonalizing the covariance matrix A,

$$A = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T \ . \tag{3.10}$$

Applying eigenvalue and eigenvector decomposition

$$\lambda_i v_i = A v_i,\ i = 1, 2, .., k.., d\ , \tag{3.11}$$

where, λ is eigenvalue and v is eigenvector of A. The eigenvalue is ordered in descending order

$$\lambda_1 > \lambda_2 > \lambda_3 > .....\lambda_d\ . \tag{3.12}$$

The corresponding eigenvector will be

$$v_1, v_2, v_3.....v_m\ . \tag{3.13}$$

After selecting k nonzero eigenvalue, eigenvector matrix is constructed as

$$v = [v_1\ v_2\ v_3.....v_k]\ ,\ k <= m\ . \tag{3.14}$$

Now input data is projected into *k* principal component axis as depicted here

$$X_{pc} = X * v\ . \tag{3.15}$$

3.5 Linear Discriminant Analysis

The most general fisher linear discriminant [33] is used for linear projection of feature to separate two or more classes. After PCA, the classification may not be efficient, because PCA does not deal with variability of features within class and between classes. To make effective and discriminative projected feature, PCA coefficients can be projected on to new LDA projection axis.

To find out class separation projection axis, there is a need to find out between-class scatter and within class variability.

The between class variable matrix can be denominated by sample variance

$$S_B = \frac{1}{c}\sum_{j=1}^{c}(m_j - m)(m_j - m)^T \quad , \tag{3.16}$$

where, c is number of classes, m is overall mean, m_i is *i-th* class mean.

Within class variance matrix can be denotes as

$$S_w = \sum_{j=1}^{c}\sum_{z_k \in w_i}(z_k - m_i)(z_k - m_i)^T \ , \tag{3.17}$$

where, z_k is *k-th* sample variable belongs to a class.

The generalized Rayleigh Coefficient is

$$J(w) = \frac{W^t S_B W}{W^t S_w W} \quad , \tag{3.18}$$

where, W matrix is for LDA coefficients. It can be characterized the generalized eigenvalue problem as

$$S_B W = \lambda S_w W \ , \tag{3.19}$$

where, λ is eigenvalue.

If S_w is singular matrix, the above equation (3.19) can be simplified as

$$S_w^{-1} S_B W = \lambda W \ , \tag{3.20}$$

where, the eigenvectors of $S_w^{-1} S_B$ will be W. The eigenvector matrix will be W_{lda},

$$W_{lda} = [W_1 \ W_2 \ W_3 ... W_k \ ], k \in Z^+ . \tag{3.21}$$

The PCA coefficients can be projected into l lower dimensional LDA projection termed by eigenvectors corresponding non-zero higher energy eigenvalues,

$$W_{lda}^{'} = [W_1 \ W_2 \ W_3 ... W_l \ ], l \in Z^+ , \tag{3.22}$$

where, $l <= k$.

The finally feature matrix, F is evaluated

$$F = (W_{lda}^{'})^T . \psi(x)_{pc} . \tag{3.23}$$

3.6 Kernel Support Vector Machine

Support Vector Machine is one of most frequently used classifier for binary classification. When the patterns are linearly separable in simplest case, linear SVM is used. When the pattern becomes non-separable, linear SVM is extended using kernel trick, mapping input pattern to higher dimensional space. The algorithm finds optimal hyperplane which maximally distant

from support vector from both classes [34]. RBF kernel SVM problem is solved by using LibSVM package [35].

The nonlinear decision function is

$$f(x) = w^T \phi(x) + w0 \ , \tag{3.24}$$

where, the kernel mapping function $\Phi : R^d \rightarrow R^f \ (f > d)$, translates the feature data to higher dimensional space so that it becomes linearly separable.

Let consider the training set $\{(x_i, y_i)\}_{i=1}^n$, where $x_i \in R^d$ training sample is and its corresponding class level is $y_i \in \{-1, +1\}$, for SVM with L1 soft margin regularization can be solved with the primal problem

$$\begin{aligned} &\min_{w,w0} \frac{1}{2} \| w \|^2 + T \sum_{i=1}^{n} \psi_i \\ &s.t. \ y_i(\langle w, \phi(x_i) \rangle + w_0) \geq 1 - \psi_i, \forall i, \\ &\xi \geq 0 \end{aligned} \tag{3.25}$$

where, T is the trade-off parameter of training error and margin ψ_{1*n} is slack vector having non-zero elements.

The Lagrange dual space optimization solution is

$$\begin{aligned} &\min_{\alpha} \ \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ &s.t. \ 0 \leq \alpha_i \leq T, \forall i; \sum_i \alpha_i y_i = 0 \end{aligned} \ , \tag{3.26}$$

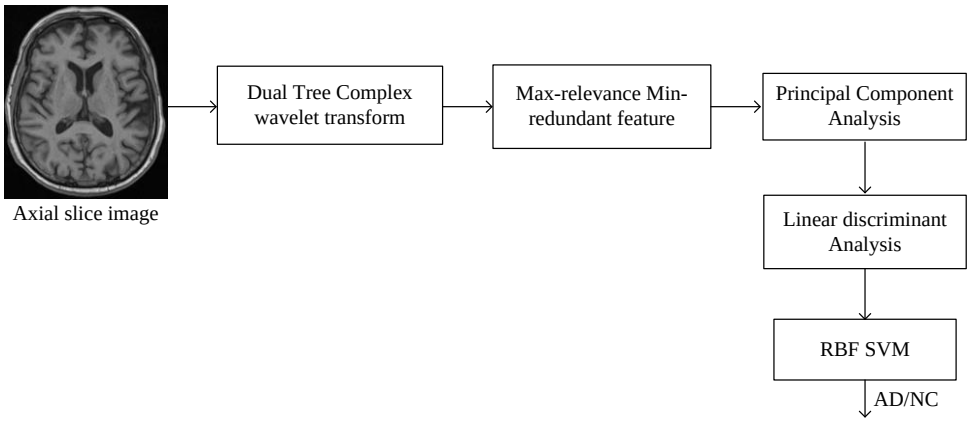


Fig. 3.3: Block diagram of dtCWT based proposed method.

where, k is the kernel matrix, α_i, α_j are the Lagrange multipliers.

Chapter 4

Performance Evaluation

4.1. Accuracy, Sensitivity, Specificity

The performance of a binary classifier can be visualized using a confusion matrix, as shown in Table 4.1. The number of examples correctly predicted by the classifier is located on the diagonal. These may be divided into true positives TP, representing correctly identified patients, and true negatives TN, representing correctly identified controls. The number of examples wrongly stratified by the classifier may be divided into false positives FP, representing controls incorrectly classified as patients, and false negatives FN, representing patients incorrectly classified as controls.

True Class	Predicted Class	
	S_1(Patients)	S_2(Controls)
S_1(Patients)	TP	FN
S_2(Controls)	FP	TN

Table 4.1: Confusion matrix for a binary classifier to distinguish between two classes (S_1, and S_2).

The accuracy measures the proportion of examples that are correctly labelled by a classifier,

$$\text{accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad , \qquad (4.1)$$

This may not be a good performance metric if the class distribution of the dataset is unbalanced.

For example, if class $C1$ is much larger than $C2$, a high accuracy value could be obtained by a classifier which labels all examples as belonging to class $C1$. Sensitivity is the rate of true positives (TP), and specificity is rate true negatives (TN). The sensitivity and specificity is defined as

$$\text{sensitivity}=\frac{\text{TP}}{\text{TP+FN}} \text{ and specificity}=\frac{\text{TN}}{\text{TN+FP}}, \tag{4.2}$$

It may provide a better assessment of the overall performance of a classifier. Sensitivity measures the proportion of correctly identified patients, and specificity measures the proportion of correctly identified controls. The balanced accuracy, which treats both classes with equal importance, may then be expressed as

$$\text{balanced accuracy}=\frac{\text{sensitivity+specificity}}{2}, \tag{4.3}$$

An ideal classifier would achieve 100% sensitivity and specificity, but in general there is a trade-off between these two measures. This can be investigated using a receiver operating characteristic (ROC) curve. As shown in Fig. 4.1, a ROC curve shows the relationship between the true positive rate (sensitivity) and false positive rate (100 − specificity) as the discrimination threshold of the binary classifier is varied. This curve could be used to select the optimal threshold for a particular application. For example, to identify patients in the earliest stages of disease, it may be desirable to select a threshold which results in high sensitivity, at the expense

of reduced specificity. The area under a ROC curve (AUC) may be interpreted as an aggregated measure of classifier performance [36].

4.2. Cross-validation

The parameters of a classifier are optimized based on the training data. An independent test set is therefore required for making a reliable assessment of the applicability of the classifier to new data. Cross-validation provides a way to measure this generalization performance when no such test data are available. One commonly used method is k-fold cross-validation, in which the data are randomly partitioned into k subsets. A single cross-validation

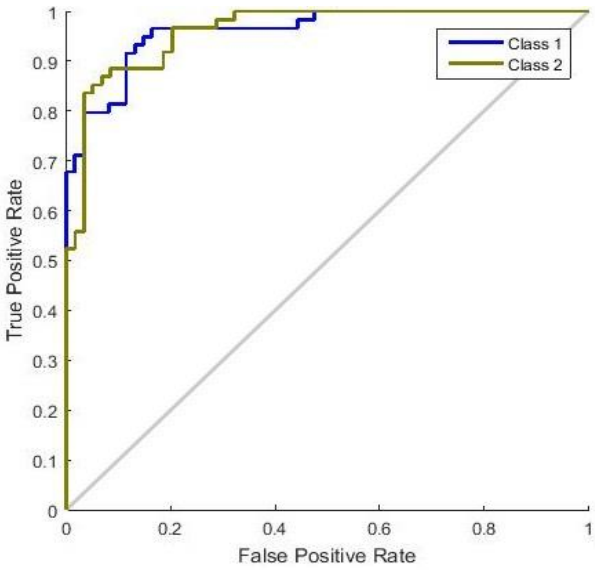


Fig. 4.1: Illustration of the ROC curve for a binary classifier. The colored solid lines show the relationship between the sensitivity and specificity as the discrimination threshold of the classifier is varied. This may be compared with the grey line of no-discrimination, and the red line depicting an ideal classifier.

fold involves using (k − 1) subsets for training the classifier, and the remaining data for testing. This process is repeated k times, such that each of the subsets is used once for testing, and the results are averaged over the folds. An alternative method is repeated random sampling, in which the dataset is randomly partitioned into training and test sets of fixed sizes. For example, a single round may involve randomly selecting 75% of the data for training, with the remaining 25% used for testing. This process can then be repeated, and the results averaged over the repetitions. Repeated random sampling has the advantage that the proportions of the training and test sets are not dependent on the number of repetitions. However, there may be some overlap between test sets, and the method also exhibits Monte Carlo variation. This means that the results will vary if the analysis is repeated using different partitions of the data. If two classes C1 and C2 are not of equal in size, the training and test sets should be selected such that they contain examples from the two classes in approximately equal proportions to the full dataset. This is known as stratified cross-validation, and has been shown to produce results with a lower variance than regular cross-validation [37].

Both the *k*-fold and repeated random sampling cross-validation methods generate a distribution of performance values which may be averaged across the folds or repetitions. The statistical significance of differences between the results of two classifiers may be assessed by performing unpaired t-tests between these distributions. In addition, permutation testing may be applied to assess whether the results of a classifier are significantly different from chance. Permutation testing involves performing cross-validation on data for which the diagnostic labels have been randomly permuted. This results in a distribution of classification results under the null hypothesis that the

classifier cannot accurately predict the clinical labels from the data. Unpaired t-tests between the distribution of observed results and that obtained from permutation testing indicate whether the observed results are significantly different from chance.

Chapter 5

Performance Analysis

5.1. Overview of Experimental Data

Data used in the preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, one can visit www.adni-info.org.

	AD	Normal
No. of subjects	86 Male-43, Female-43	86 Male-46, Female-40
Average Age	77.30	76.05
Average Education Points	14.65	15.93
MMSE	23.48	29.08

Table 5.1: Summary of subject's demographics status.

5.2. Dual Tree Complex Wavelet Transform feature

All the MR images used here are viewed using ONIS visualization tool, and extracted all the 2D slices. Some slices are subjectively selected manually where all the labels is clearly visible for wavelet based feature extraction purpose. The preprocessing and volumetric feature extraction algorithms of MR images by FreeSurfer are briefly explained in section 2.3. All MR imaging subjects details applied here are shown in Table 5.1.

5.3. Result and Discussion

In the present study, A novel approach is proposed to analyze prediction of AD based on structural MR image and its usefulness for dual disease classification. The propose method detail is sketched in Fig.3.3. To train the kernel SVM for classifying AD from NC, the training samples are prepared. The 4th level dual tree wavelet coefficients of specific scale are excerpted from selected MR slices to get comparatively optimal dimensional feature so that classifier can be trained and tested without much hassle.

The principal components of max-relevant and min-redundant dual tree wavelet coefficients are used for linear discriminant analysis. These linear discriminant coefficients are used to train kernel SVM and tested performance of accuracy, sensitivity, specificity with 10 cross validation running the program at least 20-30 times as shown details in Table 5.2. To the best of my knowledge, it is novel approach for the purpose of diagnosis which is combination of dual tree wavelet coefficients of a particular scale from selected 2D MRI image, principal components of max-relevant and min-redundant of those coefficients, linear discriminant coefficients and

Radial Basis Function kernel support vector machine for classification of the disease. The number of principal components (PCs) versus accuracy graph is shown in Fig. 5.3, and it is obvious that when the number PCs are 22, the accuracy is highest. The number of PCs is selected in similar way in all other method used here. For nonlinear mapping to higher dimensional space, Radial Basis Function (RBF) kernel is used as it performs better than polynomial kernel [38]. When principal MRMR components along with LDA are used, the performance gets higher. As explained earlier, the MRMR method selects most relevant and least redundant feature subset, and LDA does effective and discriminative projection of feature which can deal with variability within the class and between classes. Thus higher accuracy can be obtained while combining both. It has been obtained 91.807±1.15 accuracy with high sensitivity and specificity which outperforms the methods used by Zhang et al. [39], and El-Dahshan et al. [40]. The performance of these two methods is shown in Table 5.3. There is also the classification performance without using MRMR, or without LDA, or without both the methods exhibited. When the feature subset selection method is not used, the accuracy gets down. Similarly when LDA is not performed, accuracy goes down. It is depicted details in Fig. 5.1, Fig. 5.2, and Table 5.2.

There is a comparison of the method with conventional DWT based classification, and also with the proposed approaches by Zhang et al. [39], and El-Dahshan et al. [40]. The novel proposed method performs better as appeared in Table 5.3, Fig. 5.4, and Fig. 5.5. The classification performance is shown here without using max-relevant and min-redundant algorithm, or without LDA coefficients, or without both. As compared to all those method, our proposed method performs better. While principal max-relevant, min-redundant DWT coefficients are considered for classification, it performs better than principal DWT coefficient based classification. Similarly, when

those coefficients are projected into different efficient space using LDA, the performance of stratification increases. The proposed approach method achieves better accuracy with sensitivity, specificity than principal components of dtCWT and DWT coefficient based classification.

Method	Accuracy	Sensitivity	Specificity
Proposed	91.807±1.15	92.286±1.49	91.319±1.63
dtCWT+PCA+LDA+Kernel SVM	90.181±0.97	90.276±1.60	90.101±1.23
dtCWT+MRMR+PCA+Kernel SVM	84.779±0.85	84.738±0.99	84.812±1.54
dtCWT+PCA+Kernel SVM	82.735±1.24	84.4275±1.51	81.181±1.85

Table 5.2: DtCWT based classification performance of AD from HC.

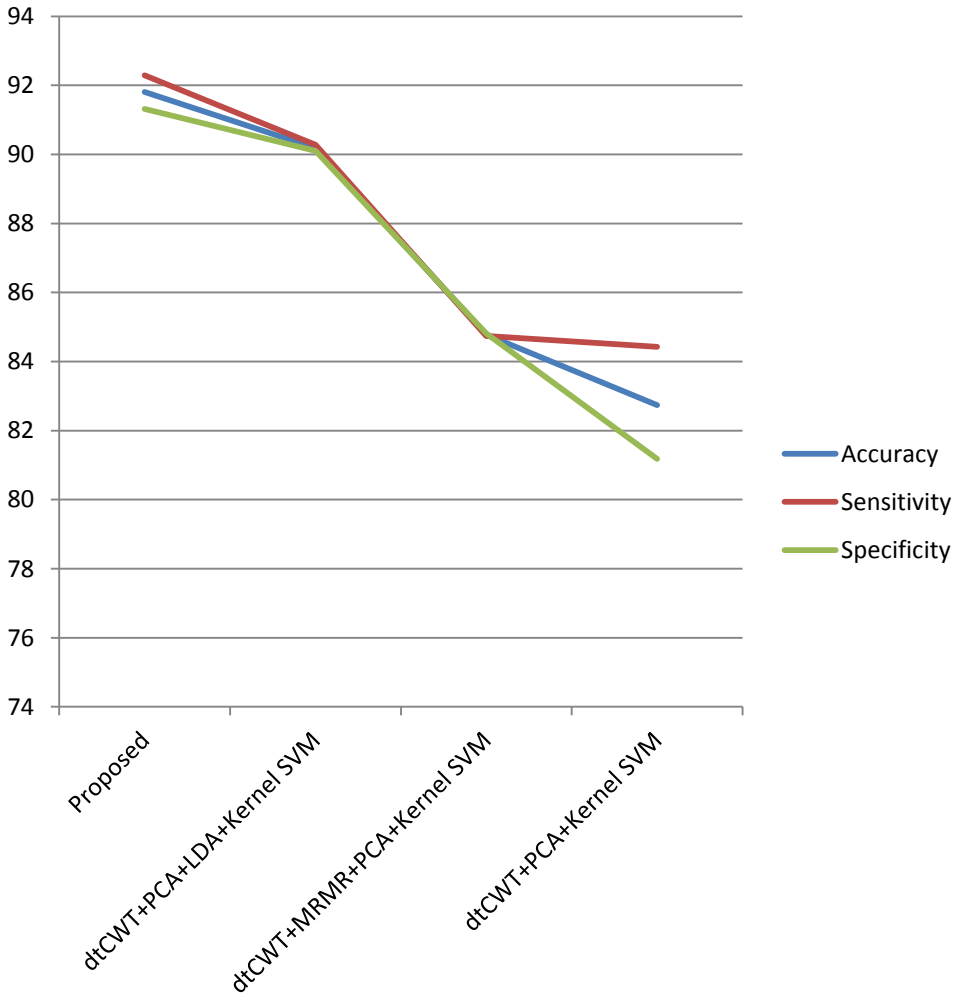


Fig. 5.1: Line chart of dtCWT based classification performance of AD from HC.

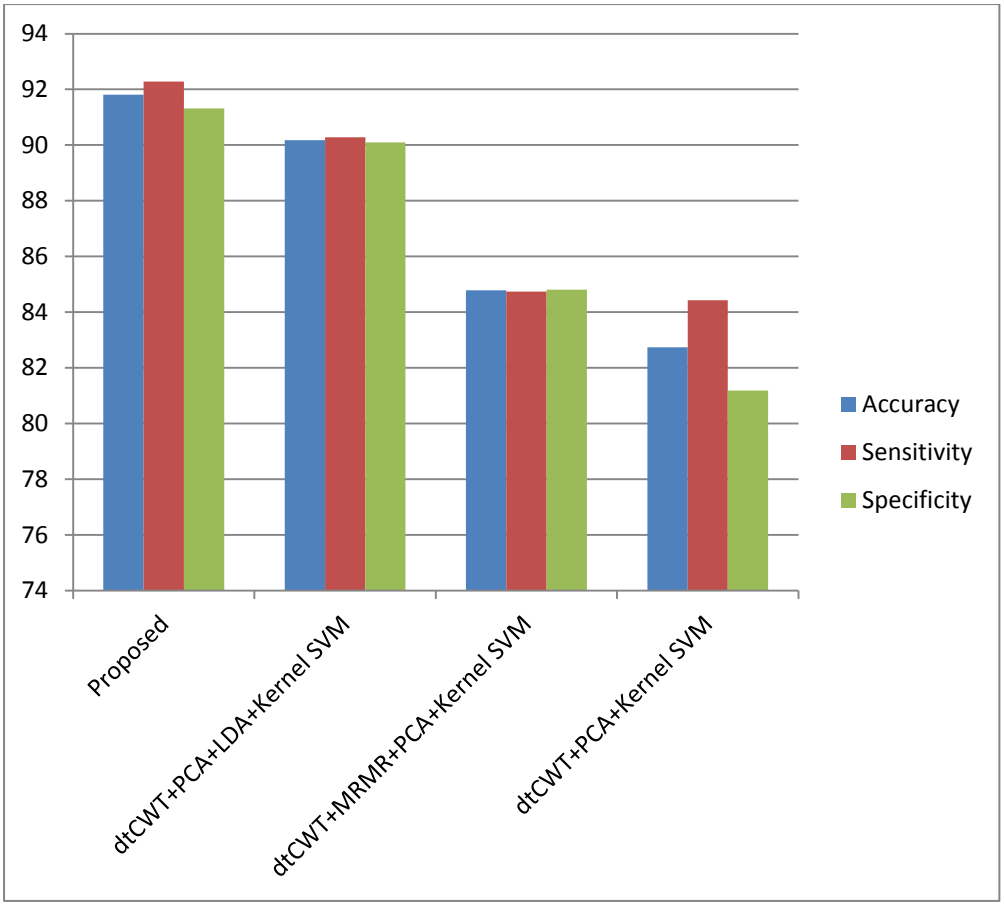


Fig. 5.2: Bar chart of dtCWT based classification performance of AD from HC.

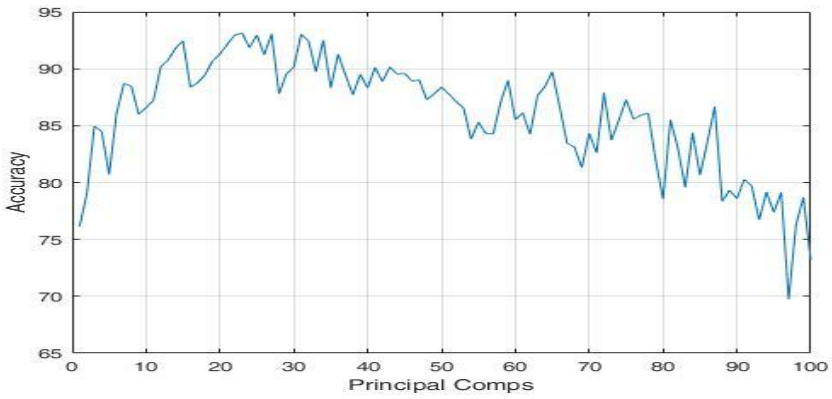


Fig. 5.3: The number of Principal Components vs Accuracy graph of proposed method.

Method	Accuracy	Sensitivity	Specificity
DWT+MRMR+PCA+LDA+ Kernel SVM	86.427±1.08	86.101±2.05	86.166±2.05
DWT+PCA+LDA+Kernel SVM	84.8219±1.13	84.648±1.54	85.100±1.32
DWT+MRMR+PCA+Kernel SVM	83.296±1.21	82.797±1.85	83.749±1.46
DWT+PCA+Kernel SVM	81.369±1.04	82.863±1.88	81.323±1.46
DWT+PCA+ANN, Zhang et al.[39]	80.05±0.72	81.538±1.41	78.974±1.09
DWT+PCA+KNN, El-Dahshan et al. [40]	79.964±1.19	78.771±2.37	81.08±1.67

Table 5.3: DWT based classification performance of AD from HC.

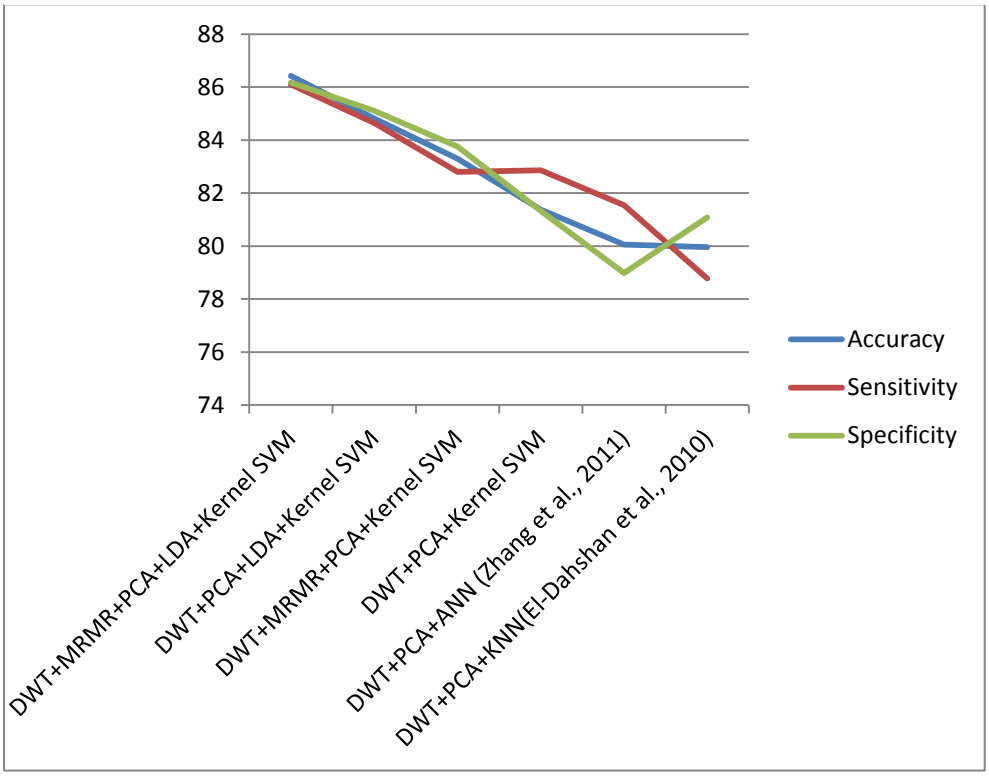


Fig. 5.4: Line chart of DWT based classification performance of AD from HC.

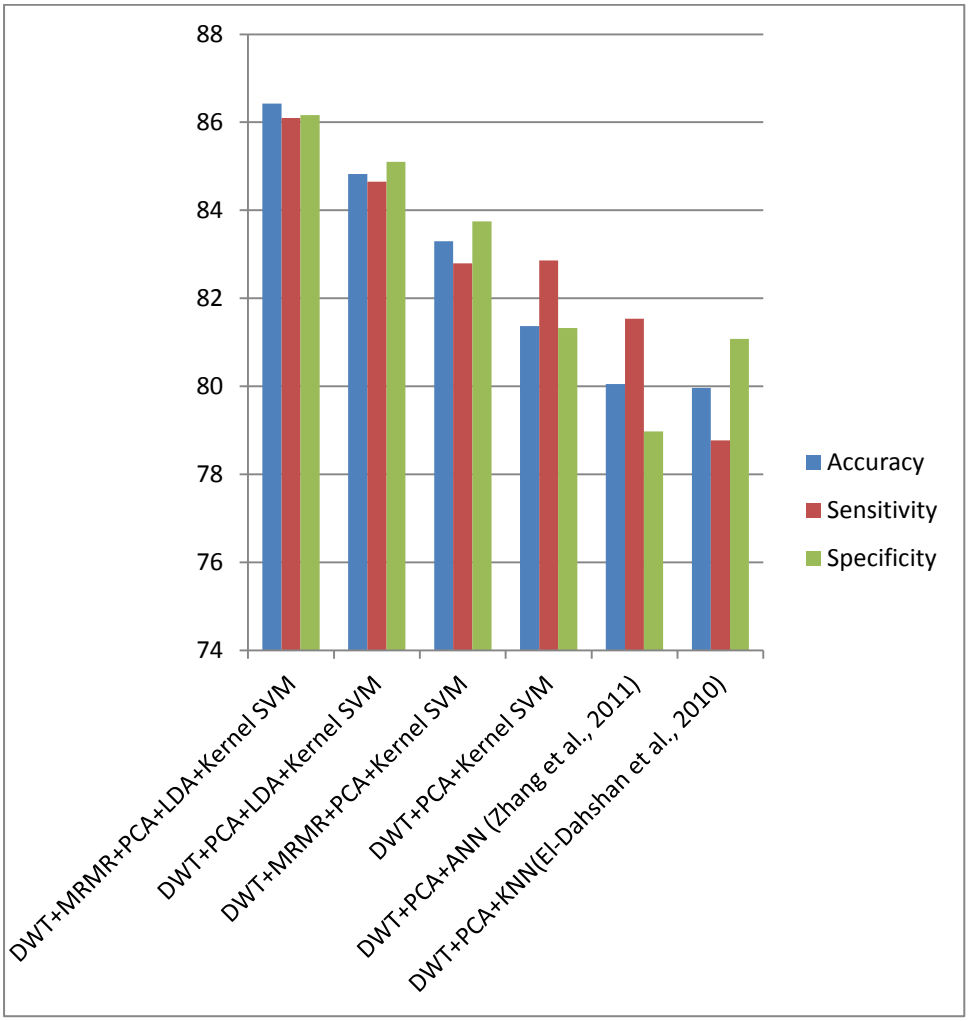


Fig. 5.5: Bar chart of DWT based classification performance of AD from HC.

The prediction accuracy has also been calculated on same dataset using FreeSurfer volumetric feature as depicted in Table 5.4, Fig. 5.6, and Fig. 5.7. The classification accuracy has been achieved up to 84.38 ±0.63 having sensitivity 85.06±1.06 and specificity 83.66±0.76. There is a similar trend observed here. When MRMR feature subset selection, or LDA coefficients are used, the performance shows an upward trend.

Method	Accuracy	Sensitivity	Specificity
MRMR+PCA+LDA+Kernel SVM	84.382±0.63	85.066±1.06	83.668±0.76
PCA+LDA+Kernel SVM	82.795±0.54	83.934±0.93	81.690±0.74
MRMR+PCA+Kernel SVM	84.067±0.67	86.749±1.20	81.39±1.20
PCA+Kernel SVM	81.629±0.71	83.629±1.15	79.627±1.47

Table 5.4: Volumetric feature (FreeSurfer) based classification performance of AD from HC.

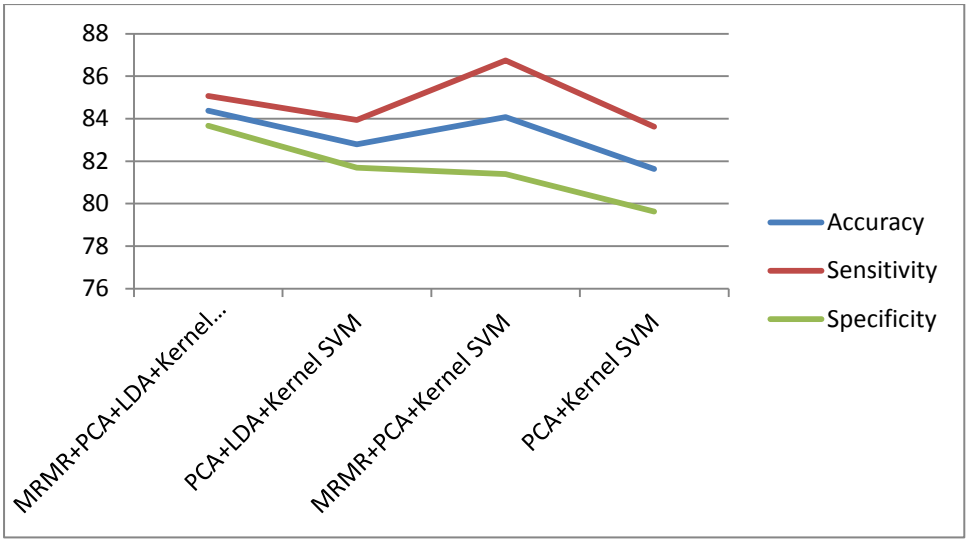


Fig. 5.6: Line chart for Volumetric feature (FreeSurfer) based classification performance of AD from HC.

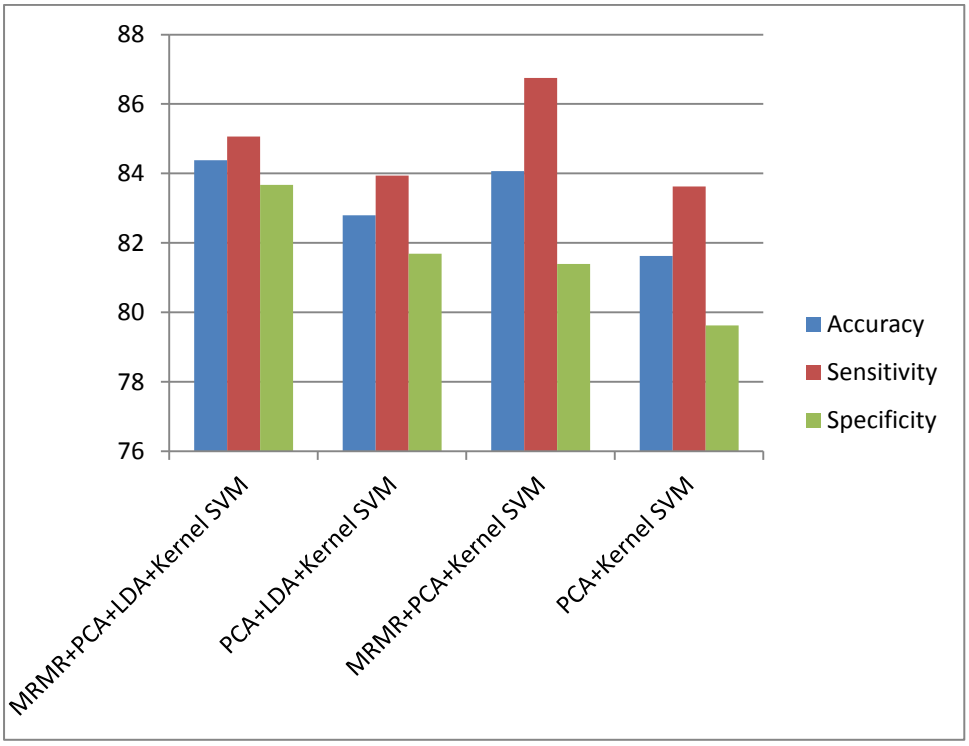


Fig. 5.7: Bar chart for Volumetric feature (FreeSurfer) based classification performance of AD from HC.

Using volumetric feature extracted by FreeSurfer, the classification accuracy has been obtained up to 85% with 88% specificity and 82% sensitivity [41]. The proposed method outperforms volumetric feature based AD prediction method, and also the method studies by Schmitter et al. [41] as illustrated in Table 5.5, Fig 5.8, and Fig 5.9.

Method	Accuracy	Sensitivity	Specificity
Proposed	91.807±1.15	92.286±1.49	91.319±1.63
Schmitter et al., 2015	85	82	88
Volumetric feature (FreeSurfer)	84.382±0.63	85.066±1.06	83.668±0.76

Table 5.5: Performance comparison of AD prediction.

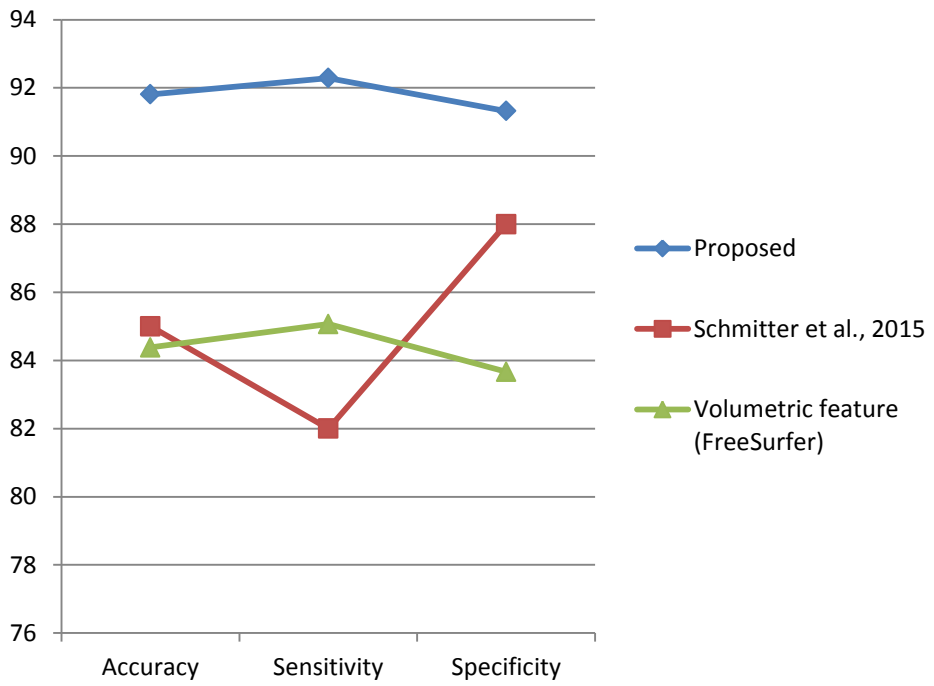


Fig. 5.8: Line chart of performance comparison of AD prediction.

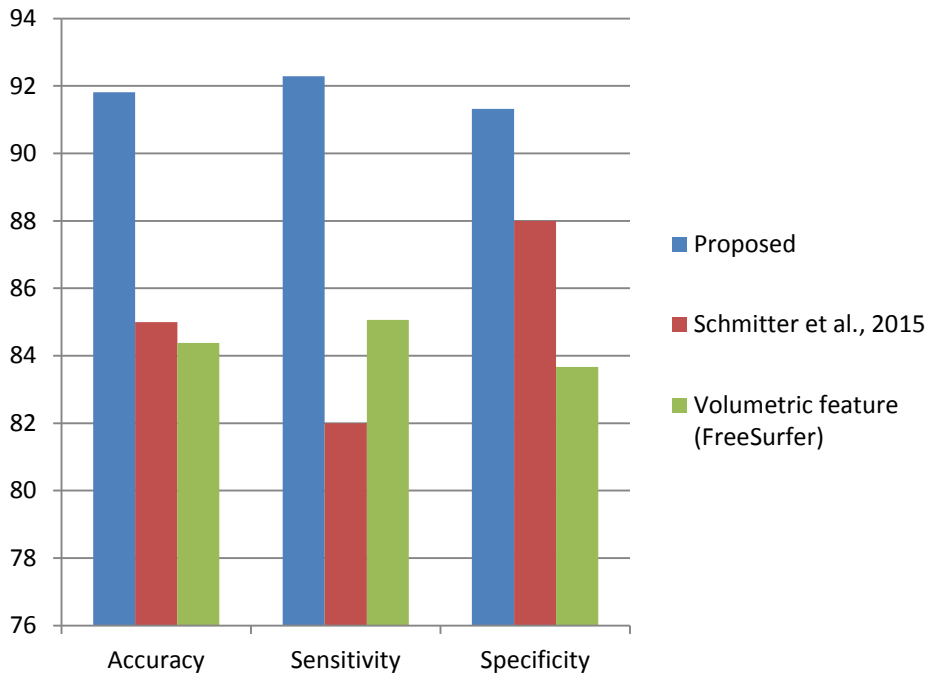


Fig. 5.9: Bar chart of performance comparison of AD prediction.

Chapter 6

Conclusion

AD prediction method plays a very significant rule in the arena of diagnosis of disease. In the modern era, plenty number of old-aged people worldwide develop AD. The irreversible neurological disorder is one major factor of death. In many countries, the amount spent for the diagnosis of the disease is also exemplary. Therefore, efficient computer aided diagnosis method is highly appealing because it beefs up the diagnosis process. Computerize aided diagnosis of a disease based on MR image has always advantage over manual perception based diagnosis done by clinicians because it may differ to other based on their experience.

In this work, dual tree complex wavelet transform (dtCWT) provides discriminative feature for selected MR slices as it considers feature details in multiple directions. The MR images are collected from publicly available ADNI dataset. Then proposed method selects feature subset from dTCWT coefficients of particular scale which is most relevant and least redundant, and then principal components are selected using PCA to get rid of curse of higher dimensionality so that classifier can be trained in lower dimensional space with hassle free. After taking variability within class and between classes into account, linear discriminant coefficients are obtained using fisher LDA. RBF Kernel SVM is trained and tested because nonlinear mapping feature to kernel space creates separable hyperplane to classify. It is presented that the feature contains substantial disease signature (disease related pattern).

To evaluate the effectiveness of the proposed method, experiments are performed and repeated 20-30 times with 10-fold cross-validation. The

performance is evaluated in terms of sensitivity, specificity and classification accuracy. It has been observed that the proposed method outperforms existing methods in terms of all three performance measures. So the proposed method has higher decision potential to assist neuro-radiologists for the diagnosis of AD from NC.

Reference

[1] N. C. Berchtold and C.W. Cotman, “Evolution in the conceptualization of dementia and Alzheimer’s disease: Greco-roman period to the 1960s,” Neurobiology of Aging, vol. 19, no. 3, pp. 173–189, 1998.

[2] L. E. Hebert, P. A. Scherr, J. L. Bienias, D. A. Bennett, and D. A. Evans, “Alzheimer disease in the US population: prevalence estimates using the 2000 census,” Archives of neurology, vol. 60, no. 8, pp. 1119–1122, 2003.

[3] Mayo Clinic Medical Information and Tools for Healthy Living. 2010 Mayo Foundation for Medical Education and Research. Web. 21 Nov. 2010.

[4] C. Hughes, L. Berg, W. Danziger, L. Coben, and R. Martin, “A new clinical scale for the staging of dementia,” Br J Psychiatry, vol. 140, pp-566-572, 1982.

[5] Todd A. Gould, and Molly Edmonds, "Discovery Health "MRI Magnets"" Discovery Health. Web. 5 Nov. 2010.

[6] "What Is A PET Scan? How Does A PET Scan Work?" Medical News Today: Health News. 22 June 2009. Web. 28 Oct. 2010.

[7] G. Emilien, C. Durlach, K. Minaker, B. Winblad, S. Gauthier, and J. Maloteaus , “Alzheimer disease: neuropsychology and pharmacology(1st ed.),” Basel: Birkhäuser Verlag, 2004.

[8] K. M. Gosche, J. A. Mortimer, C. D. Smith, et al., “Hippocampal volume as an index of Alzheimer neuropathology: findings from the Nun Study,” Neurology, vol. 58, pp. 476–1482, 2002.

[9] K. R. Gray, “Thesis: Machine learning for image-based classification of Alzheimer’s disease,” Chapter 1, pp. 35-36, 2012.

[10] H. Braak and E. Braak, “Evolution of neuronal changes in the course of Alzheimer’s disease,” Journal of Neural Transmission Supplement, vol. 53, pp. 127–140, 1998.

[11] X. Hua, A. D. Leow, N. Parikshak, S. Lee, M. C. Chiang, A. W. Toga, C. R. Jack Jr., M. W. Weiner, P. M. Thompson, and the Alzheimer’s Disease Neuroimaging Initiative, “Tensor-based morphometry as a neuroimaging biomarker for Alzheimer’s disease: an MRI study of 676 AD, MCI, and normal subjects,” Neuroimage, vol. 43, no.3, pp. 458–469, 2008.

[12] N. Schuff, N. Woerner, L. Boreta, T. Kornfield, L. M. Shaw, J. Q. Trojanowski, P. M. Thompson, C. R. Jack Jr., M. W. Weiner, and the Alzheimer’s Disease Neuroimaging Initiative, “MRI of hippocampal volume loss in early Alzheimers disease in relation to ApoE genotype and biomarkers,” Brain, vol. 132, no. 4, pp. 1067–1077, 2009.

[13] L. A. van de Pol, W. M. van der Flier, E. S. C. Korf, N. C. Fox, F. Barkhof, and P. Scheltens, “Baseline predictors of rates of hippocampal atrophy in mild cognitive impairment,” Neurology, vol. 69, no. 15, pp. 1491–1497, 2007.

[14] C. R. Jack Jr., S. D. Weigand, M. M. Shiung, S. A. Przybelski, P. C. O’Brien, J. L. Gunter, D. S. Knopman, B. F. Boeve, G. E. Smith, and R. C. Petersen, “Atrophy rates accelerate in amnestic mild cognitive impairment,” Neurology, vol. 70, no.19, pp. 1740–1752, 2008c.

[15] B. H. Ridha, J. Barnes, J. W. Bartlett, A. Godbolt, T. Pepple, M. N. Rossor, and N. C. Fox, “Tracking atrophy progression in familial Alzheimer’s disease: a serial MRI study,” Lancet Neurology, vol. 5, no. 10, pp. 828–834, 2006.

[16] B. C. Dickerson, I. Goncharova, M. P. Sullivan, C. Forchetti, R. S. Wilson, D. A. Bennett, L. A. Beckett, and L. deToledo–Morrell, “MRI-

derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease," Neurobiology of Aging, vol. 22, no. 5, pp.747–754, 2001.

[17] A. D. Leow, I. Yanovsky, N. Parikshak, X. Hua, S. Lee, A. W. Toga, C. R. Jack Jr., M. A. Bernstein, P. J. Britson, J. L. Gunter, C. P. Ward, B. Borowski, L. M. Shaw, J. Q. Trojanowski, A. S. Fleisher, D. Harvey, J. Kornak, N. Schuff, G. E. Alexander, M. W. Weiner, P. M. Thompson, and the Alzheimer's Disease Neuroimaging Initiative, "Alzheimer's Disease Neuroimaging Initiative: a one-year follow up study using tensor-based morphometry correlating degenerative rates, biomarkers and cognition," Neuroimage, vol. 45, no. 3, pp. 645–655, 2009.

[18] E. Saltiel, "Memantine (Namenda)." MedicineNet. 2 Nov. 2003. Web. 12 Nov. 2010.

[19] E. A. Mellon, et al., "Sodium MR Imaging Detection of Mild Alzheimer's Disease: Preliminary Study," American Journal of Neuroradiology, vol. 30, no. 5, pp. 978-984, 2009.

[20] Nordberg, "PET Studies and Cholinergic Therapy in Alzheimer's Disease," Rev Neurol 4th ser.155, pp. 53-63, Oct. 1999.

[21] University of South Florida Health, "MRI Brain Scans Accurate In Early Diagnosis Of Alzheimer's Disease," ScienceDaily, 23 December 2008. Web. 11 November 2010.

[22] K. Bethune, "Thesis: Diagnosis and Treatment of Alzheimer's Disease: Current Challenges," Chapter 1, pp. 12-13, 2010.

[23] B. Fischl, "*FreeSurfer*," NeuroImage, vol. 62, pp. 774–781, 2012.

[24] R. S. Desikan, F. Segonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," NeuroImage, vol. 31, pp. 968-980, 2006.

[25] Fischl et al., “Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain,” Neuron, vol. 33, pp. 341-355, 2002.

[26] Fischl et al., “Automatically Parcellating the Human Cerebral Cortex,” Cerebral Cortex, vol. 14, no. 1, pp. 11-22, 2004.

[27] Klein and J. Tourville, “101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol,” Frontiers in Neuroscience, vol. 6, pp. 171, 2012.

[28] R. C. Gonzalez and R.E. Woods, “Digital Image Processing, second ed.,” Pearson Education, Ch. Wavelet and Multiresolution Processing, pp. 349–408, 2004.

[29] J. Koenderink, “The structure of images,” Biol. Cybern., vol. 50, no. 5, pp. 363–370, 1984.

[30] A. Barri, A. Dooms, and P. Schelkens, “The near shift-invariance of the dual-tree complex wavelet transform revisited,” Journal of Mathematical Analysis and Applications, vol. 389, no. 2, pp. 1303–1314, 2012.

[31] H. Peng, F. Long, and C. Ding, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy,” IEEE transactions on pattern analysis and machine intelligence, vol. 27, no. 8, pp. 1226-1238, 2005.

[32] H. Andersen, D. M. Gash, and M. J. Avison, “Principal component analysis of the dynamic response measured by fMRI: a generalized linear systems framework,” Journal of Magnetic Resonance Imaging, vol. 17, pp. 795–815, 1999.

[33] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” Annals of Eugenics, vol. 7, pp. 179–188, 1936.

[34] Burges, “Tutorial on support vector machine for pattern recognition,” Data Mining Knowl. Discov, vol. 2, no. 2, pp. 121-167, 1998.

[35] C. Chang and C. -J. Lin, “LIBSVM: a library for support vector machines,” 2001.

[36] P. Flach, J. Hernandez-Orallo, and C. Ferri, “A coherent interpretation of AUC as a measure of aggregated classification performance,” In Proceedings of ACM 28th Annual International Conference on Machine Learning, pp. 657–664, 2011.

[37] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” In Proceedings of 14th International Joint Conference on Artificial Intelligence, vol. 2, pp. 1137–1143, 1995.

[38] 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.

[39] Y. Zhang, Z. Dong, L. Wu, and S. Wang, “A hybrid method for MRI brain image classification,” Exp. Syst. Appl., vol. 38, pp. 10049–10053, 2011.

[40] El-Dahshan, T. Hosney, and A.B.M. Salem, “Hybrid intelligent techniques for MRI brain images classification Digital,” Signal Process., vol. 20, no. 2, pp. 433–441, 2010.

[41] D. Schmitter, A. Roche, B. Maréchal, et al., “An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer’s disease,” J. NeuroImage: Clinical, vol. **7**, pp. 7–17, 2015.

List of Publications

Saruar Alam and Goo-Rak Kwon, "Alzheimer Disease classification by combining SRAN, L-SVM, PCA, VBM," The Journal of Korean institute of next generation computing, Vol. 9, No, 37-46, Aug. 2015.

Saruar Alam, Dibash Basukala, GyoungTae Ha, Chuntao Wang, Goo-Rak Kwon, "Deep Model for Improved Classification of AD/NC Patients," 1st International Conference on Next Generation Computing 2016, Jan. 2016.

Dibash Basukala, **Saruar Alam**, Debesh Jha, Sang-Woong Lee, Nguyen Van Han, Jae-Young Pyun, and Goo-Rak Kwon, "An Advanced Face Detection and Recognition using Combination of KPCA, LDA, and Multiclass L-SVM," 1st International Conference on Next Generation Computing 2016, Jan. 2016.

Saruar Alam, Goo-Rak Kwon, Moonsoo Kang, and Jae-Young Pyun, "Performance of classification based on PCA, Linear SVM, and Multi-kernel SVM," The Eighth International Conference on Ubiquitous and Future Networks 2016, July. 2016. (Accepted)

Saruar Alam and Goo-Rak Kwon, "Alzheimer disease classification using KPCA, LDA, Multi-kernel learning SVM," International Journal of Imaging Systems and Technology, 2016. (Submitted)

Saruar Alam and Goo-Rak Kwon, "Alzheimer disease prediction based on minimum redundancy maximum relevance complex dual tree wavelet

principal coefficients and kernel SVM.," The Journal of Korean institute of next generation computing, 2016. (Submitted)

Acknowledgement

It is pleasure for me to present this dissertation. I would like to express my sincere appreciation to all those who played a part in the accomplishment of my dissertation, supported and guided me throughout the study of my MS degree.

First of all, I would like to express my deep gratitude and special appreciation to my advisor, Prof. Goo-Rak Kwon for his guidance, encouragement, constant support, and all the invaluable lessons learned. Under his supervision, I have learnt to identify and approach cutting edge research trends and problems. Under his thoughtful guidance, the exposure I have achieved during my research would be a real treasure for my life.

I would like to convey my heartfelt gratitude to committee members Prof. Jae-Young Pyun, and Prof. Kim Young-Sik for providing constructive criticism, and expert and thoughtful suggestions.

I sincerely thank my all lab members of Digital Media Computing Lab especially Dr. Ramesh Kumar Lama and the friends who had cooperated and encouraged throughout my MS course

I express my deepest gratitude to my parents and family for their love, support, and encouragement throughout of my entire life. Their moral and emotional support has made the way apparent to pursue my MS degree.

Lastly, I would like to acknowledge the financial assistance provided by the Center for Dementia and Alzheimer disease, Chosun University.