August 2017
Master's Degree Thesis

Alzheimer's Disease Classification using DTCWT, PCA and Feed-forward Neural Network in MRI

Graduate School of Chosun University

Department of Information and Communication Engineering

Debesh Jha

Alzheimer's Disease Classification Using DTCWT, PCA and Feed-forward Neural Network in MRI

MRI 영상에서 DTCWT 와 PCA 그리고 Feed-forward Neural Network 알고리듬을 이용한 알츠하이머 병변 분류

August 25, 2017

Graduate School of Chosun University

Department of Information and Communication Engineering

Debesh Jha

Alzheimer's Disease Classification Using DTCWT, PCA and Feed-forward Neural Network in MRI

Advisor: Prof. Goo-Rak Kwon

A thesis submitted in partial fulfillment of the

requirements for a master's degree

April 2017

Graduate School of Chosun University

Department of Information and Communication

Engineering

Debesh Jha

자 데베쉬의 석사학위논문을 인준함

위원장 조선대학교 교수 변 재영 (인)

위 원 조선대학교 교수 이 범식 (인)

위 원 조선대학교 교수 권 구락 (인)

2017년 5월

조선대학교 대학원

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AD	Alzheimer's Disease
CAD	computer aided diagnosis
CDR	clinical dementia rating
CWT	complex wavelet transform
DWT	discrete wavelet transform
HC	healthy controls
PET	position emission tomography
fMRI	functional magnetic resonance imaging
SPECT	single-photon emission computed tomography
SNR	signal-to-noise ratio
ANN	artificial neural network
KNN	k-nearest neighbor
BRC	brain region cluster
IG	information gain
FFT	fast Fourier transform
DBM	deformation-based morphometry
TJM	trace of Jacobian matrix
VBM	voxel-based morphometry
GEODAN	Geodesic Anisotropy
PEC	pearson's correlation
LS-SVM	least square support vector machine
SCV	stratified cross validation
GARCH	generalizing autoregressive conditional heteroscedasticity
PCNN	pulse-coupled neural network
SVD	singular value decomposition
DWPT	discrete wavelet packet transform

FNN	feed-forward neural network
DTCWT	dual-tree complex wavelet transform
EB	expectation-maximization
SCG	scale conjugate gradient
DF	displacement field
PCA	principal component analysis
CV	cross-validation
ROI	region of interest
DWT	discrete wavelet transform
SSD	single slice based detection
OASIS	open access series of imaging studies
MMSE	mini-mental state examination
CDR	clinical dementia rating
SVM	support vector machine
BPNN	back-propagation neural network
LDA	linear discriminant analysis
CNN	convolutional neural network
PSO	particle swarm optimization
BBO	bibliography based optimization
RF	random forest
WTT	Welch’s test
RBF	radial basis function

ABSTRACT

Alzheimer's Disease Classification Using DTCWT, PCA, and Feed-forward Neural Network in MRI

Debesh Jha
Advisor: Prof. Goo-Rak Kwon
Dept. of Information and Communication Engineering
Chosun University

Alzheimer's disease (AD), the most familiar form of dementia, is a neurodegenerative disorder of the brain that causes memory loss to the elderly people. Magnetic resonance imaging (MRI) of the brain provides comprehensive diagnostic information for diagnosis. The error-free diagnosis of Alzheimer's disease (AD) from Healthy control (HC) at the early stage is a major concern, because the knowledge of the severity and the development risks allows the subjects to take precautionary measures before irretrievable brain damages are shaped. Recently, there have been great interests for computer-aided diagnosis for magnetic resonance image (MRI) classification. However, identifying the distinctions between Alzheimer′s brain data and healthy brain data in older adults is challenging due to highly similar brain patterns and image intensities. Recently, cutting-edge feature extraction technologies have been very quickly expanding into numerous fields, incorporating medical image analysis. In this paper, we proposed dual tree complex wavelet transform (DTCWT) for extracting features from the image. The dimensionality of feature vector is reduced by using principal component analysis (PCA). The reduced feature is sent to feed-forward neural network (FNN) to distinguish MR images into AD and NC.

These proposed and implemented pipelines, which demonstrate an improvement in classification output when compared to other studies, resulted in high and reproducible accuracy rates of 90.06±0.1 % with sensitivity of 92.00±0.40 %, a specificity of 87.78±0.40 % and a precision of 89.6 ± .03 % with 10-fold cross validation.

한 글 요 약

MRI 영상에서 DTCWT 와 PCA 그리고 Feed-forward Neural Network 알고리듬을 이용한 알츠하이머 병변 분류

데베쉬 자
지도교수 : 권구락
정보통신공학과
조선대학교

알츠하이머병(Alzheimer's disease)은 노인들에게 기억 상실을 일으키는 뇌의 퇴행성 신경 장애로서 치매의 가장 친숙한 형태이다. 뇌의 자기공명영상(MRI)은 포괄적인 진단 정보를 제공한다. 건강관리를 통한 알츠하이머병의 오류 없는 초기 진단은 주요 관심사이며, 이는 심각한 발달 장애 및 뇌 손상으로 인한 지능 저하 전에 피험자가 사전 예방 조치를 취할 수 있게 해주기 때문이다.최근에는, 자기공명영상 분류를 위한 컴퓨터 보조 진단에 많은 관심이 일어나고 있다. 그러나 고령자의 경우에는 알츠하이머병의 뇌 데이터와 건강한 뇌 데이터의 차이를 확인할 때, 매우 유사한 뇌 패턴과 이미지 강도로 인해 많은 어려움이 있다. 최근에는 최첨단 특징 추출 기술이 의료 영상 분석을 포함하여, 다양한 분야로 빠르게 확대되고 있다.본 논문에서는 이미지로부터 특징을 추출하기 위해 DTCWT(Dual Tree Complex Wavelet Transform)를 제안한다. DTCWT를 수행한 대역의

계수들을 주성분 분석(PCA)을 사용하여 특징 벡터의 차원을 감소시킨다. 감소된 특징은 피드 포워드 뉴럴 네트워크(FNN)로 전송되어 MR 영상을 알츠하이머병과 일반영상으로 구분한다.다른 연구와 비교했을 때 분류 결과가 개선되었음을 입증하는 제안 및 구현된 파이프 라인은 민감도 92.00±0.40%, 특이성 87.78±0.40% 및 정확도 90.06±0.1%의 높은 재현성 정확도를 나타낸다. 10배 교차 유효성 확인으로는 89.6±0.03%가 실험결과에 의해 증명된다.

CHAPTER 1

1. Introduction

1.1 Overview and motivation

Alzheimer disease (AD) is an irremediable neurodegenerative disorder that causes dementia in the elderly people globally. It leads to progressive deterioration of memory, cognition function and behavior of the patient. It is the dominant healthcare concern in the near future. It has been predicted that the pervasiveness of AD will double in the next 2 decades, by which time and 1 out of 85 people will be affected with the disease by 2050. The prevalence of AD differs among various factors, including age, co-morbidities, genetics, and educational status. There is no method to definitely diagnose AD without performing an autopsy. There is no cure of AD, nevertheless promising study and development for early diagnosis and treatment is underway. Therefore, it is imperative to determine effective treatments towards not only the symptoms, but also the reason of the disease. The accurate early diagnosis AD, especially identifying the risk of progression of MCI to AD, affords the AD patients awareness of the severity and allows them to take preventative measures, e.g., lifestyle changing and medications. To clinically diagnose the AD subject at the primitive stage, MRI is influential brain imaging technique that provides correct information about the shape and volume of the brain as compared to different modalities. One of the most well-known dataset utilized for early detection of AD progression is Open Access Series of Imaging Studies (OASIS) which are aimed at making the dataset freely available to the scientific community and hope to facilitate new discoveries in basic and clinical neuroscience. A computer-aided diagnosis (CAD) platform is gaining popularity because it overcomes the disadvantage of manual methods.

1.2 Objectives.

The objective of this thesis is to study the brain with the Alzheimer's disease and develop a support system that could be useful to the medical doctors for detection of AD from healthy controls.

1.3 Contribution

According to best of my understanding, a noble approach is utilized in this work. For this method, 'whole brain analysis' method is utilized since it takes all the voxels in the brain as whole. It is not essential to segment the brain as earlier, and it does not require any biomarker for the classification purpose. I have considered multiple slices for each patient so that information gained is more consistent, reliable, and accurate. In hospitals, multiple slice based detection is utilized because of its inexpensiveness. The dual tree complex wavelet transform is utilized on selected 2D MR image slices for extraction of features. The dual tree complex wavelet transform has a property of shift invariance and directional selectivity. A principal component analysis (PCA) is utilized for the dimensionality reduction of the DTCWT coefficients. The output from the PCA acts as an input to the feed-forward neural network which classifies the input image as healthy or abnormal.

The proposed method is implemented using the 32-bit Matlab 2015b environment on Intel(R) Core (TM) i3-2120, with a processing speed of 3.30 GHz and 2GB of RAM running Microsoft Windows 7. I have compared and analyzed the potential of the proposed system in terms of Accuracy, sensitivity, specificity and precision and compared our result with the recent classification technique.

1.4 Structure of the thesis

The thesis is organized as follows. Chapter 2 offers background knowledge about Alzheimer's disease, problem statement and progress in diagnosis, Detection technique, Magnetic resonance imaging, and Image processing and normalization. Chapter 3 includes Literature review. Chapter 4 contains the brief description about the feature extraction, feature reduction and classification technique along with performance estimation and cross-validation technique. The Proposed method, overview of the experimental data and Results and discussions are analyzed in Chapter 5, Chapter 6 and Chapter 7 respectively. Finally the conclusion and the future research are represented in Chapter 8.

CHAPTER 2

Background

2.1 Alzheimer's disease

Alzheimer's disease (AD) is the most familiar form of dementia that causes problems with memory, thinking, behavior, and other intellectual abilities significant enough to interfere with day-to-day life [1]. Symptoms usually develop slowly and worsen over time, ultimately becoming serious enough to interfere with daily life [2]. Alzheimer's disease reports for the majority of (60% to 80%) of dementia cases. In spite of Alzheimer's disease progresses differently for every individual, there are numerous common symptoms. In the primitive stages, the most familiar symptom is problem in remembering recent events, well known as short term memory loss. The prevalence of AD fluctuates among many different factors, comprising age, co-morbidities, genetics, and education level. There is no way to absolutely diagnose AD without accomplishing an autopsy. There is no cure for AD, nevertheless encouraging research and development for early detection and medical care is underway.

History

Alzheimer's disease was detected in 1906 by Alois Alzheimer's, a German neurologist and psychiatrist [3]. The disease was firstly noticed in a 51-year-old lady named Auguste D. Her family took her to Dr. Alzheimer's in 1901 after observing changes in her nature and behavior. The family reported problems with memory, complication speaking, and impaired comprehension. Dr. Alzheimer's later reported Auguste as having an aggressive form of dementia, manifesting in memory, language and behavioral deficits. Dr. Alzheimer distinguished many abnormal symptoms, including complication

with speech, agitation, and confusion. He observed her for almost five years, until she died in 1906. Following her death, Dr. Alzheimer's carried out an autopsy, during which he found noticeable shrinkage of the cerebral cortex, fatty accumulations in blood vessels, and atrophied brain cells. He found neurofibrillary tangles and senile plaques, which have turn out as indicative of AD. The condition was first argued in medical literature in 1907 and titled after Alzheimer in 1910.

Early-dementia

This mild stage generally lasts 2 to 4 years. In this level, family and friends may start to realize that there has a deteriorating in the patient's cognitive intelligence. Common symptoms at this stage involve:

- inconvenience retraining new information
- Trouble with problem solving or decision making. Subject may start to have difficulty managing finances or alternative instrumental activities of day-to-day living.
- Personality changes. The individual may start to withdraw socially or exhibit lack of motivation.
- Difficulty signifying thoughts
- Misplacing belongings or getting lost. The subject may have problem navigating in common surroundings.
- Having greater difficulty executing tasks

Moderate Alzheimer's disease

This is longest stage of the disease that lasts for 2 to 4 years. Subjects often experience increased problems with memory and may need assistance with activities of daily living. Symptoms commonly reported during this stage involve:

- Increasingly poor judgement and confusion.
- Difficulty accomplishing complex tasks, containing many of the instrumental activities of everyday life, such as directing finances, marketing, planning, and organization.
- Greater memory loss. Subjects may start forgetting their detail personal history.
- Notable personality changes. The individual may become detached from social interactions and develops unusual suspense to the caregivers.
- confusion regarding where they are and what day it is
- Alteration in sleep patterns, such as sleeping throughout the day and becoming anxious at night

Severe Alzheimer's disease

In this last stage of the disease, cognitive volume continues to decreases and physical ability is harshly impacted. The final stage can last for 1 and 3 years. Because of family's deteriorating ability regarding the patients concern; this stage often seen in long term care facility placement and nursing homes. Common symptoms seen in this stage are:

- Loss of capability to exchange information. The patient may however speak short phrases, but are unable to carryout coherent conversation.
- Dependence on others for personal concern, for example eating, bathing, walking, dressing, and toileting.
- Incapacity to function physically. The individual may be incapable to walk or sit independently. Muscles may grow rigid and swallowing can ultimately be impaired.

2.2 Detection Techniques

Neuroimaging is promising area of study for detecting AD. In addition, to diagnose an AD sufferer clinically at a primitive disease stage, many imaging biomarkers must be identified using different imaging modalities-, such as (MRI) [4], position emission tomography (PET) [5], functional Magnetic resonance imaging (fMRI) [6], single-photon emission computed tomography (SPECT) [7], and magnetic resonance spectral imaging (MRSI) [8], etc. Each scan involves a unique method and detects specific anatomy and abnormalities in the brain. Brain imaging technique is not currently a benchmark of AD testing; nevertheless recent clinical studies have revealed promising results which may change the procedure utilized by physicians to detect the disease.

2.3 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is most popular imaging technique that provides comprehensive diagnostic information for diagnosis [9]. It is essential because it is non-invasive brain imaging technique that provides higher-quality statistics about the structure and volume of the brain. It produces high quality and greater contrast image of the anatomical structures along with functional images of various organs [10]. The rich information content of the high quality MRI makes it better tool for both clinical diagnosis as well as biomedical research [11]. The features such as soft tissue differentiation, high spatial resolution and ability to identify a tiny irregularity inside the brain make it superior than other existing technique [12-14]. This imaging method does not cause any radiation harm to the tissues as it is not utilizing any unfavorable ionizing radiation to the subjects. Moreover, the diagnostic use of MRI has been tremendously improved due to the automated and precise labeling of MR images, which performs an

important role in identifying AD in related patients from healthy control (HC) and elderly controls.

MRI is mainly used to diagnose different types of disease such as strokes, tumors, bleeding, injury, blood vessel diseases or infections and multiple sclerosis etc. Classification of normal/pathological brain status from MRIs is essential in clinical medicine since MRI focuses on soft tissue anatomy and generates an enormous information set and details about the subject's brain situations.

2.4 MRI Pros and Cons

When we consider the influence of this imaging method, there exist both pros and cons. Potential advantage of choosing this approach are that it is pain-free and can detect very small abnormalities without the radiation hazard of an X-ray. The resulting image also has tremendous spatial resolution. Nevertheless, this process is very costly and insurance company may not refund for it. Due to very strong magnets utilized in the MRI scanner, metal is forbidden anywhere nearby the scanner. This means that subjects with metal fragments inside their physique cannot go through an MRI. This comprises people with pacemakers, artificial joints, dental fillings, braces, etc. In order for a favorable MRI scan, the subject must lie perfectly still. This is a difficulty for many patients, mainly children.

2.5 Accuracy of MRI

Florida Alzheimer's Research Center conducted a research and found that MRI scans are successful in detecting the brain atrophy noticed in AD. They gathered brain scans for 260 subjects, some with mild cognitive impairment, and rest with possible AD, and a control group of aged adults with no memory deterioration. The investigators were able to match the brain scans

with the appropriate group of patients deepening on the quantity of atrophy in the mid-brain. Few scans exhibited brain atrophy before any sign were present, suggesting that this technique would be adequate for early diagnosis of the disease. Sample transverse slices form MR images of healthy individuals and AD patients are depicted in Figure 10. These image clues that AD subjects usually show proof of cortical atrophy, and enlarged ventricles in resemblance with healthy controls.

2.6 Image preprocessing and normalization

For each patient, each scanning session involves the MR of three or four T1-weighted image scans. In order to add the signal-to-noise ratio (SNR), all indicated MRI scans with the identical protocol of the same individual are motion-corrected and spatially co-registered, to the Talairach coordinate space to produce an averaged image, and then are brain-masked. The motion-correction recorded the 3D images of all scans, and then developed an average 3D image in initial acquisition space. Also, the scans are then resampled to 1 mm × 1 mm × 1mm. The obtained image is converted from acquisition space to Talairach coordinate space. Lastly, the brain extraction is achieved.

MRIcro software (http://www.cabiatl.com/mricro/mricro/) is utilized and imported the image from the backup folder and then extracted 2-D MRI slices of the each subjects. In this paper, I only choose 32 important center slices from each subject manually based on our experience. These slices are used for preprocessing. The reason behind picking center slice from all slices is it retains more relevant information about the brain tissues as compared to earlier slices and later slices in group of (1-256) slices. The direction of the slice possibly may be sagittal, coronal, or axial. In this research, axial direction is chosen by knowledge. The same process is applied to all the

subjects (126 including both AD and HC). All images are in PNG format and the dimension of the slices are 176×208. The image is resized to 256×256 before used for the further processing.

CHAPTER 3

Literature review

Scholars have proposed numerous methods to extract various features. Chaplot et al. [13] used the approximation coefficients acquired by discrete wavelet transform (DWT). Maitra and Chatterjee [14] used the slantlet transform that is an enhanced version of DWT. The feature vector of each image was produced by examining the magnitudes of slantlet transform outputs equivalent to six spatial positions that were selected according to a specific logic. El-Dahshan et al. [15] extracted the approximation and detail coefficients utilizing 3-level DWT. They utilized classifier based on artificial neural network (ANN) and K-nearest neighbors. Plant et al. [16] used brain region cluster (BRC). They suggested to apply information gain (IG) to rate the charm of a voxel, and used clustering algorithm to recognize groups of adjacent voxels with a huge discriminatory power. Zhang et al. [17] particularly decomposed the image into 3-level of decomposition, utilized its approximation coefficients, and then applied PCA to decrease the features. Ramasamy and Anandhakumar [18] used fast Fourier transform (FFT) as features. Sarita et al. [19] proposed a new feature of wavelet-entropy, and used digital wavelet transform to extract features then applied PCA to reduce the feature space. Savio and Grana [20] suggested using deformation-based morphometry (DBM) approach, and suggested five features as Jackbian map, modulated GM (MGM), trace of Jacobian matrix (TJM), significance of the displacement field, and Geodesic Anisotropy (GEODAN). In addition, they suggested to use the Pearson's correlation (PEC), Bhattacharyya distance (BD), and WTT to measure the significance of voxel site. Das et al. [21] suggested combining ripplet transform, PCA for dimensionality reduction and least-square SVM (LS-SVM) for classification, and the 5×5 stratified

cross validation (SCV) offered high accuracy. Kalbkhani et al. [22] designed the detail coefficients of 2-level DWT by generalizing autoregressive conditional heteroscedasticity (GARCH) statistical model, and the parameters of GARCH model were treated as the primary feature vector. Zhang et al. [23] used an under sampling (US), method on the volumetric image, accompanied by singular value decomposition (SVD) to choose features. El-Dahshan et al. [24] proposed to add a preprocessing technique that utilized pulse-coupled neural network (PCNN) for segmentation of the image. Zhou et al. [25] applied wavelet-entropy as the feature space. Zhang et al. [26] employed discrete wavelet packet transform (DWPT), and harnessed Tsallis entropy to get features from DWPT coefficients. Yang et al. [27] selected wavelet-entropy as the features. From the literature review, the results with DWT based features were considered as excellent. Debesh Jha et al. [28] used discrete wavelet transform with PCA and ensemble classifier and achieved promising result in classifying Alzheimer's disease and dementia. Later on, Debesh Jha et al. [29] utilized deep learning algorithm along with sparse autoencoder, scale conjugate gradient and softmax output layer and achieved significant improvement in this results. In this study, I have suggested using a novel feature of dual-tree complex wavelet transform (DTCWT), PCA and feed-forward neural network (FNN) for classification of the MR image into healthy control (HC) or Alzheimer's disease (AD).

CHAPTER 4

Wavelet Transform

4.1 2D-Discrete wavelet Transform

It is an effective implementation of the WT utilizing the dyadic scales and positions [30]. The basic fundamentals of DWT are introduced as follows. Let $x(t)$ be a square integral function. The continuous WT of the signal $x(t)$ relative to a real-valued wavelet $\psi(t)$ is defined as [31]

$$W(a,\tau) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{a}}\psi^{*}(\frac{t-\tau}{a})\,dt \tag{1}$$

where, $W(a,\tau)$ is the WT, τ interpret the function across $x(t)$, and the variable α is the dilation factor (both real and positive numbers). Here, the asterisk (*) indicates the complex conjugate.

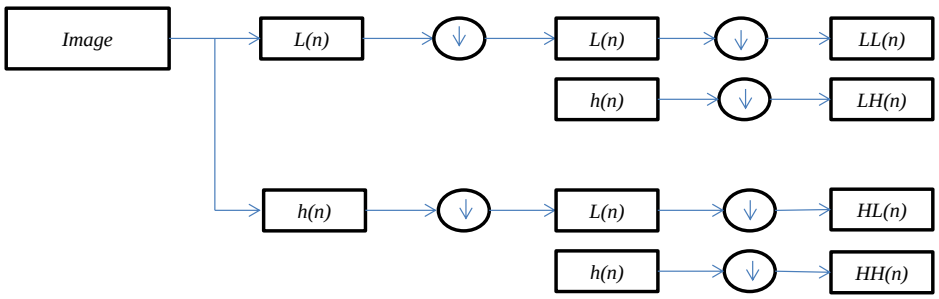
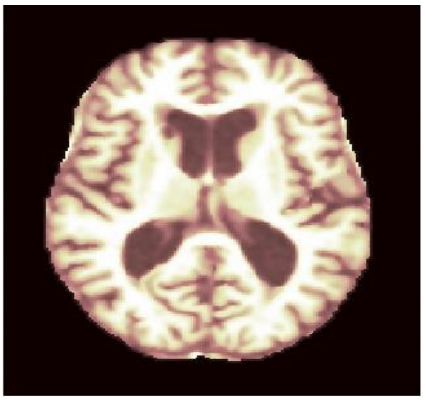


Figure1: A brain image and its wavelet coefficient at level-3 decomposition

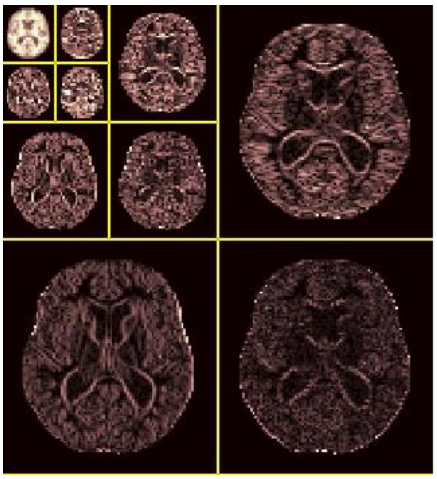
Equation (1) can be discretized by controlling α and τ to a discrete lattice ($a = 2^{j}$ and $\tau = 2^{j}k$) to provide the DWT which can be signified as follows:

$$cA_{j,k}(n) = DS\left[\sum_{n} x(n)l_{j}^{*}(n-2^{j}k)\right]$$

$$cD_{j,k}(\mathrm{n}) = \mathrm{DS}\left[\sum_{n} x(\mathrm{n})\mathrm{h}^{*}_{j}(\mathrm{n}-2^{j}\mathrm{k})\right] \tag{2}$$



(a)



(b)

Figure 2: (a) Transverse slice MR normal brain image (b) Decomposition at level 3 using 2D-discrete wavelet transform along with Haar filter

Here, $cA_{j,k}$ and $cD_{j,k}$ refer to the coefficients of the approximation components and detail components, respectively. l(n) and h(n) stands for the low-pass and the high-pass filter, respectively. j and k denotes the wavelet scale and translation factors, correspondingly. The DS operator means down sampling. The approximation component holds low-frequency components of the image, whereas the detailed components contain high frequency. A brain image and its wavelet coefficient at level-3 decomposition is demonstrated in Figure 1.The principal feature of DWT is multiscale representation of function. By utilizing wavelets, the given function can be determined at various levels of resolution [32]. Figure 2 shows the schematic diagram of 2-level 2D-DWT. There are various types of wavelets such as Daubechies, symplets, Biorthogonal, Haar etc. Haar wavelet is one of the oldest, simplest and important wavelets in the discrete wavelet family. Because of its simplicity it is often preferred [33].

4.2 Dual-tree complex wavelet Transform

The Dual Tree Complex wavelet Transform (DTCWT) is a modified version of the traditional DWT. To help boost the directional selectivity impaired by DWT, DTCWT is proposed. The traditional DWT is shift variant on account of the decimation operation used in the transform. As a consequence, a small shift in the input signal can create a very dissimilar set of wavelet coefficients formed at the output. It utilizes two real DWTs processing input data in parallel [34]. The first DWT symbolizes the real component of the transform, whereas the second DWT depicts the imaginary component together forming a complex transform. Figure 4 shows the implementation of DTCWT utilizing its two wavelet filter banks in parallel.

The DTCWT provides a solution for “shift-invariant problems” as well as for “directional selectivity in two or more dimensions,” which are both shortcomings of the ordinary DWT [35]. The DTCWT obtains directional selectivity by utilizing wavelets that are approximately analytic. It has the

ability to produce a total of six directionally discriminating sub-bands oriented in the (± 15, ± 45, and ± 75 directions), for both the real (R) and imaginary (I) parts. Figure 2 illustrates the DTCWT. Let $h_i(n)$ and $g_i(n)$ are the filters in the first stage as in Figure 2. Let, the new kth stage response of the first filter bank by $H_{new}^{(k)}(e^{jw})$ and second filter bank by $H_{new}^{'(k)}(e^{jw})$, we now have following result as a corollary of Lemma 1.

Collary1: Suppose we are provided CQF pairs $\{h_o(n), h_1(n)\}, \{h_o^{'}(n), h_1^{'}(n)\}$.

For $k > 1$

$$H_{new}^{(k)}(e^{jw}) = H\{H_{new}^{'(k)}(e^{jw})\} \tag{3}$$

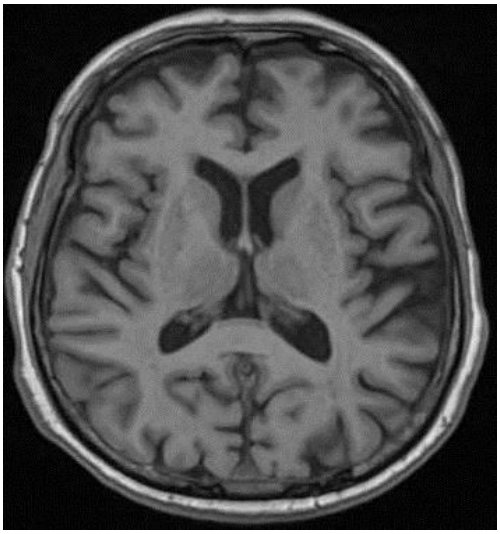
If and only if

$$h_0^{'(1)}(n) = h_0^{(1)}(n-1). \tag{4}$$

A 2-D image $f(x,y)$ is decomposed by 2-D DTCWT over a series of dilations and translations of a complicated scaling function and six complex wavelet functions $\varphi_{j,l}^{\theta}$, i.e.,

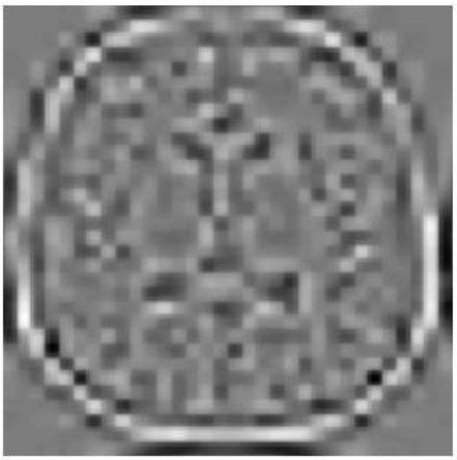
$$f(x,y) = \sum_{l \in Z^2} s_{jo,l} \phi_{jo,l}(x,y) + \sum_{\theta \in \Theta} \sum_{j \geq jo} \sum_{l \in z^2} c_{j,l}^{\theta} \varphi_{j,l}^{\theta}(x,y) \tag{5}$$

where, $\theta \in \Theta = \{\pm 15^0, \pm 45^0, \pm 75^0\}$ gives the directionality of the complex wavelet function. This is to say that, the decomposition of $f(x,y)$ by utilizing the DTCWT creates one complex valued low-pass subband and six complex valued high-pass subband at every level of decomposition, where every high-pass subbands corresponds to one particular direction θ.



(a)

(b)



(c)

Figure 3: Transverse slices of MR images (a) Original image taken for reconstruction, (b) An image reconstructed by DTCWT; reconstructed error is 0.7326 (c) An image reconstructed after utilizing DWT, reconstructed error obtained is 0.8812.

The comparison of sample slice after DTCWT reconstruction and DWT reconstruction is demonstrated in Figure 3.

A study was carried out in [36] to compare the DTCWT's directional selectivity to that of the DWT. The simulation results showed that, the edges detected by the DTCWT had clear contours, and nearly all directions could be detected clearly and perfectly. However, the edges detected by the DWT were discontinuous, and only horizontal and vertical edges could be successfully detected. The results verify the effectiveness of the DTWCT over the DWT. By utilizing DTCWT, I have extracted DTCWT coefficients from the preprocessed images. The additional features include the information about the demographics of the patients such as age, gender, handedness, education, SES, and clinical examination etc. The handedness features are not included in the work since all the patients are right-handedness.

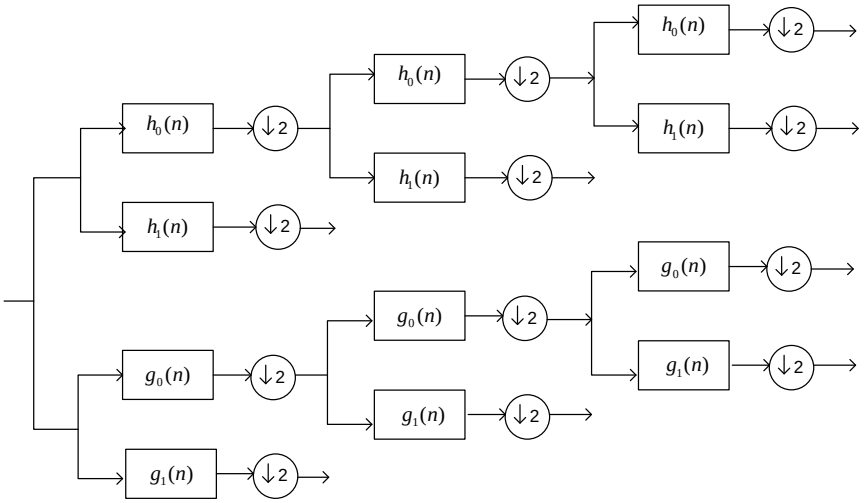


Figure 4: The DTCWT is implemented utilizing two wavelet filter banks functioning in parallel.

4.3 Principal component analysis

Let X be an input text file (X : matrix of dimensions $M \times N$).

Accomplish the following steps:

Step1. Estimate the empirical mean: $u[\mathrm{m}] = \frac{1}{N} \sum_{n=1}^{N} X[\mathrm{m,n}]$.

Step 2. Compute the deviations from the mean and save the data in the matrix $B[\mathrm{M} \times \mathrm{N}] : \mathrm{B} = \mathrm{X} - \mathrm{u.h}$, here, h is a $1 \times N$ row vector of all 1's:

$h[\mathrm{n}] = 1 \text{for } n = 1, ..., N.$

Step 3. Obtain the covariance matrix $C : C = \frac{1}{N} B.B^{*}$.

Step 4. Get the eigenvectors and eigenvalues of the covariance matrix $V^{-1}CV = D : V$ - the eigenvectors matrix; D-the diagonal matrix of eigenvalues of $C, D[\mathrm{p,q}] = \lambda_m$ for $p = q = m$ is the *mth* eigenvalues of the covariance matrix C.

Step5. Rearrange Eigen vectors and eigenvalues: $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq .....\lambda_N$.

Step 6. Selecting components and developing a feature vector: save the first L columns or V as the $M \times L$ matrix W ,

$W[\mathrm{p,q}] = \mathrm{V[p,q]}, \text{for } p = 1, ...M, q = 1....., L \text{ where } 1 \leq L \leq M.$

Step 7. Obtaining the fresh data set: The eigenvectors with the leading eigenvalues are forecasted into space, this projection appears in a vector depicted by fewer dimension $(L < M)$

Figure 5: PCA algorithm

The coefficient from the DTCWT enlarges the dimensionality of feature space that makes the classification job more complicated. Additionally, it leads to excessive computational overhead and enormous memory storage. As a result, it is essential to minimize the dimension of the feature set and get the significant features to boost the classification result.

Since last two decades, a method called PCA has earned much more attention for data visualization and reduction of dimensionality. It systematically projects the initial input data to a lower dimensional space, well-known as principal subspace through an orthogonal transformation while preserving most of the data variations with a low computational cost [37]. For a stated set of likely correlated variables, this transformation outcomes in a set of values of linearly uncorrelated variables, known as principal component (PCs). All of the steps to implement PCA are demonstrated in Figure 5.

The additional information on PCA and its implementations can be viewed in literature [38, 39]. Let us consider a set of data. PCA is employed to find a linear lower-dimensional reduction of the data set. In this case, the variance of the constructed data is preserved. PCA limits the feature vectors to the component it selects, which leads to an effective classification algorithm. The main idea behind implementing PCA is reduction of the dimensionality of the DTCWT coefficients, which results in more adequate and accurate classification.

The following algorithm is utilized to obtain the principal components from the input matrix and finally fed to the feed-forward artificial neural network. Now, the input matrix possesses only these principal components. Hence, the size of the matrix is reduced. Therefore, feature extraction is done in two steps: DTCWT extract the wavelet coefficients and essential coefficients are later selected by the PCA as described in Figure 5.

4.4 Probabilistic Principal Component Analysis

The PPCA algorithm suggested by Tipping and Bishop [40-42] is based on estimating the principal axes when any input vector has one or more missing values. The PPCA reduces the high-dimensional matrix to a lower dimensional representation by relating p-dimensional observation vector y to a k-dimensional latent (or unobserved) variable x, regarded as normal with zero mean and covariance $I(\mathrm{k})$. Moreover, PPCA depends on an isotropic error model. The relationship can be established as

$$y^T = W * x^T + \mu + \varepsilon \tag{6}$$

where, y denotes the row vector of observed variable, ε denotes isotropic error term and x is the row vector of latent variables. The error term, ε, is Gaussian with zero mean and covariance $v * I(\mathrm{k})$ where v is the residual variance. To make residual variance greater than 0, the value of k should be smaller than the rank. Standard principal component with v equals 0 is the limiting condition of PPCA. The observed variables, y, are provisionally independent for the given values of the latent variables, x. Therefore, the correlation between the observation variables is explained by the latent variables and the error justifies the variability unique to the y_i. The dimension of the matrix W is $p \times k$ and it relates both latent and observation variables. The vector μ allows the model to acquire nonzero mean. PPCA considers the values to be missing and arbitrary over the dataset.

From this model,

$$y: \ N(\mu, \mathrm{W} * \mathrm{W}^T + \mathrm{v} * \mathrm{I}(\mathrm{k})) \tag{7}$$

Given that the solution of W and V cannot be determined analytically, expectation-maximization (EM) algorithm are used for iterative maximization of the corresponding log-likelihood function. The EM algorithm considers missing values as additional latent variables. The columns of W span the solution subspace. PPCA then yields the orthonormal coefficients. Because the computational cost is higher PPCA utilize for the dimensionality reduction. The advantage of PPCA over PCA is its computationally efficient.

4.5 Feed-forward Neural Networks

Feed-forward neural networks (FNN) are broadly used in pattern classification because they do not need any information regarding the probability distribution or a priori probabilities of distinct classes. Neural networks (NN) harness power from their densely parallel structure, and their ability to acquire information from experience. As a result, they can be utilized for accurate classification of input data into different classes, provided that they are pre-trained. The architecture of multilayer feed-forward neural network is shown in Figure 6.

Three factors need to be considered in designing an ANN for a definite application. (i) The topology of the network; (ii) The training algorithm and (iii) The neuron activation function. A network may have many layers of neurons, and its complete architecture may possess either a feedforward or a back propagation structure.

A single-hidden-layer back propagation NN with sigmoid neurons in its hidden layer is chosen. Similarly, linear neurons are selected for the output layer. The training vector is provided to the NN, which is trained in batch mode [43]. The NN is a two-layer network, and its configuration is

$N_I \times N_H \times N_0$, N_I represents the input neurons, N_H is the hidden layer, and N_0, indicates that the brain under observation is either HC or AD.

4.5.1 Training method.

Mathematicians have already proven that a conjugate gradient (CG) algorithm, probing along conjugate gradient directions, produces a faster convergence than the steepest descent directions do. Among CG algorithm, the scaled conjugate gradient (SCG) method is the most powerful. Thus, the SCG is utilized to train our network.

Let ω_1 and ω_2 be the connection weight matrix linking the input layer and hidden layer, and the hidden layer and the output layer, respectively. Later, the training process can be deduced reported by the following equations to improve these weighted values that can be divided into four subsequent steps [44].

1. The calculation of the outputs of every neuron in the hidden layer is done by:

$$y_j = f_H\left(\sum_{i=1}^{N_I} \omega_1(i,j)\, x_i\right) \qquad j = 1, 2, \ldots\ldots, N_H \tag{8}$$

Here x_i stands for the ith input value, σ stands for the y_j output of the hidden layer, and f_H is referred as the activation function of hidden layer, commonly a sigmoid function as observed:

$$f_H(x) = \frac{1}{1+\exp(-x)} \tag{9}$$

2. The outputs of every neurons in the output layer are stated as follows:

$$O_k = f_0(\sum_{j=1}^{N_H} w_2(\mathrm{j},\mathrm{k})\, \mathrm{y}_j)\ \mathrm{k} = 1, 2, ..., \mathrm{N}_s \tag{10}$$

Here f_o represents the activation function of output layer that is usually a line function. At first, all weights are accredited with random values, and amended by the delta rule on the basis to the learning samples.

3. The error is articulated as the MSE of the distinction among output and target value [45].

$$E_l = mse(\sum_{k=1}^{N_O} O_k - T_k)\ \ l = 1, 2, ..., N_S \tag{11}$$

where T_k depict the kth value of the genuine labels which is already well-known to users, and N_S denotes the number of samples [46].

Let's consider that there are N_S samples, therefore, the fitness value can be written as

$$F(\omega) = \sum_{l=1}^{N_S} E_l \tag{12}$$

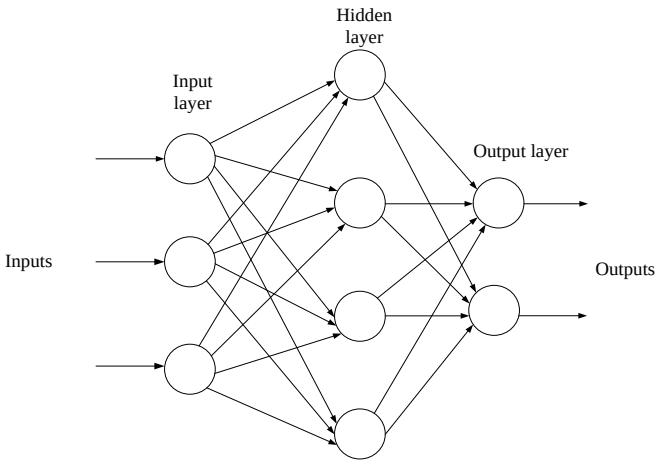


Figure 6: Architecture of a multilayer feedforward neural network.

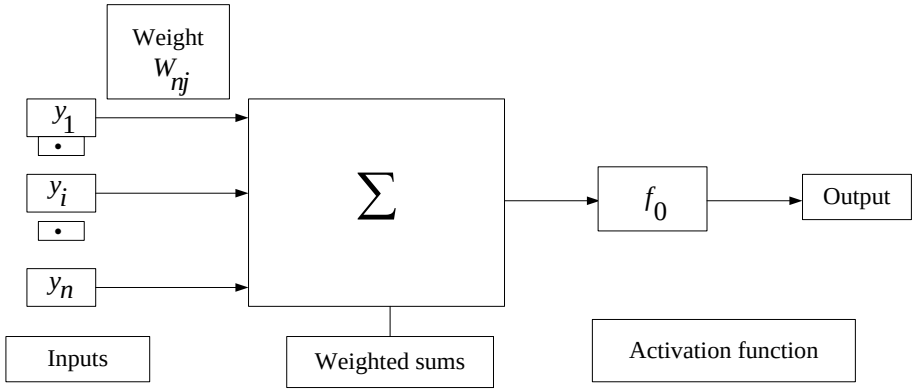
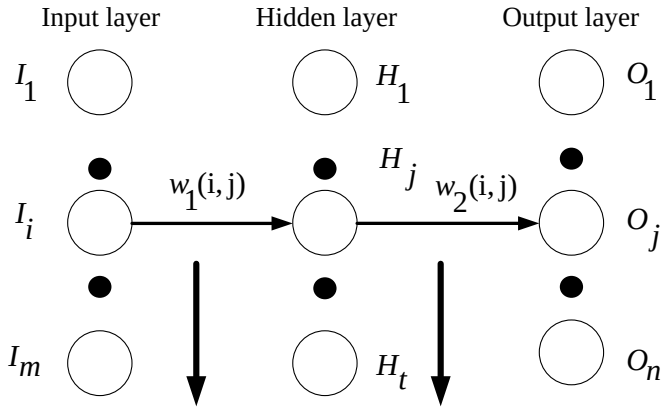


Figure 7: Hidden or output layer j: The input j are outputs from the previous layers. These are multiplied by their equivalent weights to form a weighted sum.



$$w_1 = \begin{bmatrix} w_1(1,1) & w_1(1,2) & \dots & w_1(1,H_t) \\ w_1(2,1) & w_1(2,1) & \dots & w_1(2,H_t) \\ \dots & \dots & \dots & \dots \\ w_1(I_m,1) & w_1(I_m,2) & & w_1(I_m,H_t) \end{bmatrix} \qquad w_2 = \begin{bmatrix} w_2(1,1) & w_2(1,2) & \dots & w_2(1,O_n) \\ w_2(2,1) & w_2(2,2) & \dots & w_2(2,O_n) \\ \dots & \dots & \dots & \dots \\ w_2(H_t,1) & w_2(H_t,2) & & w_2(H_t,O_n) \end{bmatrix}$$

(a) (b)

Figure 8: Connection weight matrix between (a) input layer and hidden layer and (b) hidden layer and output layer.

A non-linear activation function is used to the net input. [The inputs to input j are labeled $y_1, y_{2,\ldots\ldots\ldots} y_n$ .If unit j were in the first hidden layer, then these inputs would resemble to the input tuple [$(I_1, I_2, I_3, \ldots \mathrm{I}_n)$.] Where ω designate the vectorization of the (ω_1, ω_2) .

The aim is to minimize the fitness function $F(\mathrm{w})$, viz., force the output values of every sample appropriate to equivalent target values. The hidden layer or the output layer j is depicted in the Figure 7.

The inputs, weighted sum, and activation function of output layer is shown in the above Figure. The connection weight between the input layer and hidden layer and hidden layer and output layer is shown in the Figure 8. The connection weights can also be represented in the matrix form known as connection weight matrix.

4.7 performance estimation

To ensure unbiased testing, the data must be split-up into two sets: a training and test set. Additionally, it is generally suggested to choose a bigger training set in order to improve classifier convergence. Indeed, the performance of the learned classifier rely on how the initial data are partitioned into training and test set, and, most severely, on their size. In other words, the Connection weight between (a) input layer and hidden layer and (b) hidden layer and output more instances we allow for test, the fewer model remain for training, and thus the classifier becomes the less accurate. In contrast, a classifier which can explain one set of data strongly might not necessarily speculate to other sets of data even if it is from the same distribution. The complication known as “the curse of dimensionality” may occur. One way to overcome this issue is the utilization of “cross-validation.”

This process provides the accurate estimation of the classifier performance. The main goal is to recognize the best parameters (e.g., C, d, and σ) which can accurately forecast unknown data.

4.8 Cross-validation

Cross-validation (CV) is a model estimation technique used to check out the performance of machine learning algorithms in making predictions on unknown or new datasets. This can be achieved by portioning a dataset and utilizing a subset for training an algorithm and the remaining data is utilized for testing. As CV does not utilize all of the data to develop a model, it is commonly utilized for preventing overfitting during training. Each round of CV includes randomly portioning the original dataset into a training set and a testing set. The training set is utilized to train a supervised learning algorithm and the testing set is utilized to measure the performance. This procedure is repeated several times and the average CV error is utilized as a performance indicator. Common CV method includes:

- K-fold: Divides data into k randomly selected subsets (or folds) of approximately equal size. One subset is utilized to justify the model trained utilizing the remaining subsets. This procedure is repeated k times such that each subject is used exactly once for validation.
- Leave out: Divides data utilizing the k-fold method where k is equivalent to the total number of inspection in the data. This is also well-known as leave-one-out cross-validation technique.

In this paper, we have employed K-fold cross validation method to validate our results.

CHAPTER 5

Proposed method

Earlier, majority of diagnosis work was accomplished manually or semi-manually for measuring a priori region of interest (ROI) of MRI, based on the reality that subjects with AD experience more cerebral atrophy when compared to HCs. Most of this ROI-based examination focused on the contracting of the cortex and hippocampus, and amplified ventricles. Nevertheless, ROI-based approaches are not practicable in hospitals because of few shortcomings: (1) ROI technique needs a priori data and expert knowledge. (2) The manual diagnosis accuracy is dependent on the knowledge of physicians and interpreter. (3) The interaction among the voxels was troublesome to enforce. (4) It was essential to explore other potential areas that may be linked to AD. (5) Automatic segmentation of ROI was not beneficial in practice, and investigator needed to segment the brain using hand. Therefore, automated methods can assist physician in diagnosing diseases from images such as those produced by MRI, for which many slices are extracted from the tissues and long periods of may be necessary for the evaluation of the images.

The aim of this article is to present an automated approach for diagnosing AD by using the “whole brain analysis” method. It has achieved popularity since it takes all voxels in the brain as a whole. It is not essential to segment the brain as earlier, and it does not require any biomarker for the classification purpose. The main drawback is dimensionality that can be resolved through high-speed computers, which is comparably inexpensive. The whole-brain investigation laboriously relies on true computation, and it can only be finished by computer researcher after physician assist label the input data as either AD or HC. Usually, the whole-brain inspection labels the

entire brain as a ROI, where two stages are involved namely, feature extraction and classification.

Scholars have presented different methods to extract effective features for detection of AD and other types of pathological brain disease. Additionally, different classification models and methods survive; nevertheless, not all of them are suitable for processing of MR brain images. Based on latest literature, two drawbacks are found with the previous work. First, the discrete wavelet transform (DWT) is usually utilized for feature extraction. The DWT has better directional selectivity in horizontal, vertical and diagonal directions and has better image representation than Fourier transform but its major drawbacks are that it has poor directionality, is sensitive to shifts, and lacks phase information. Second, most of state-of-the-art mechanisms consider only single slice based detection (SSD) per patient. The obtained slices may not contain the foci of the disease.

To tackle these problems, two improvements are suggested. First of all, DTCWT is proposed that possesses attractive properties for image processing, including, shift invariance and high directionality. Secondly, unlike previous studies, multiple slices are considered for each patient so that information gained is more consistent, reliable, and accurate. In hospitals, multiple slice based detection is utilized because of its inexpensiveness. Research has clearly showed that the DTCWT is more suitable than the traditional wavelet domain for feature extraction.

The proposed CAD system for the AD detection is described in this section. The architecture of the suggested system is shown in Figure 9. As specified in the introduction, the aim of the proposed system is to boost the system performance during diagnosis of AD. In order to reach this goal, a cascade system is designed. The proposed method has four major components,

namely, image preprocessing and normalization, dual-tree complex wavelet transform, PCA, and feed-forward neural network which are used for feature extraction, feature reduction, and classification respectively. In the first stage, preprocessing of the image is done including resizing and slice selection. In the second stage, dual-tree complex wavelet technique is employed for feature extraction. In the third stage, all the approximation coefficients from the sub-bands of 5-level from dual-tree complex wavelet transform are sent to PCA for dimensionality reduction. In the fourth stage, the different new feature set are fed to the feed-forward classifier for classification, meanwhile the optimal number of principal component is selected which can obtain the most accurate results.

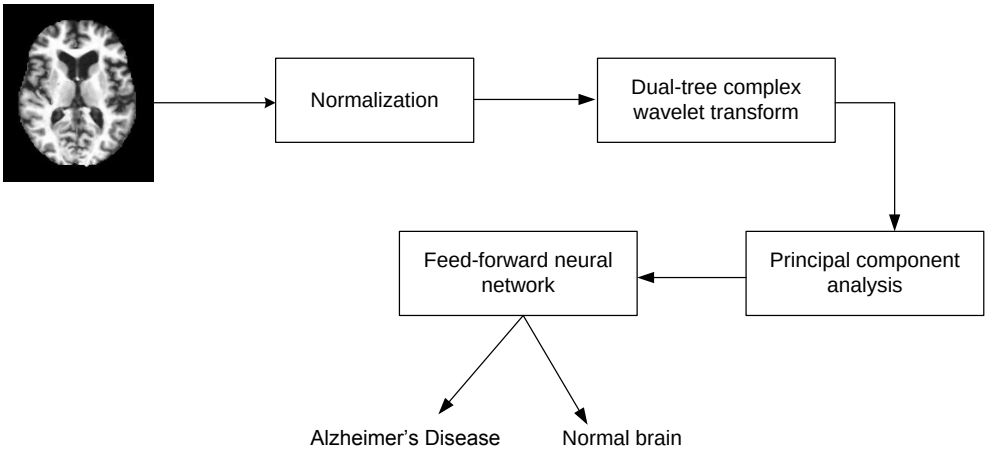


Figure 9: Illustration of proposed system.

Finally the proposed model conducts the diagnostic tasks using optimal parameters of the feed-forward classifier. The pseudocode of the suggested method is described step wise in Table 1. All of the preprocessing methods are used to obtain good results. To show effectiveness of our proposed system, I have evaluated performance measures including accuracy,

sensitivity, specificity, precision and bar- plot for comparison of the proposed algorithm with the existing systems.

Table 1: Pseudocode of the proposed system

Step 1: Import.

a. Import the OASIS dataset.

b. Ensure MRI as Normal or abnormal brain

Step 2: Resample the image into 256×256.

Step 3: Compute 5-level DTCWT on the preprocessed images.

Step 4: Perform PCA on the obtained matrix. The selected number of principal component (PC) should preserve at least 90% of total variances.

Step 5: Train feed-forward artificial neural network by taking input as reduced set of feature vectors and their corresponding class labels.

Step 6: Evaluation

a. Obtain the confusion matrix.

b. Calculate the classification accuracy and other essential parameters.

CHAPTER 6

Overview of the experimental Data

In our study, the dataset is downloaded from Open Access Series of Imaging Studies (OASIS). OASIS is a project, for compiling and sharing MRI data sets of the brain to make such data accessible to the scientific community. The data are accessible at http://www.oasis-brains.org. A sample of the MR brain image is shown in Figure 10.

OASIS provides two types of data: cross-sectional and longitudinal MRI data. In this study, a cross-sectional MRI data is used because I have aimed to develop an automatic system for detecting AD, which would not require longitudinal data that had been gathered from AD patients over long periods of time.

The dataset consists of 416 subjects whose ages are between 18 and 96. In our study, I have considered 28 AD patients and 98 healthy individuals. Table 4 shows statistical information about the subjects included in the experiment. Only right-handed subjects are included in the study, consisting of both men and women. The exclusion criterion is patients less than 60 years of age or any of their reports are missing. The unbalanced data may cause difficulty in future recognition; I have fine-tuned the cost matrix to resolve this issue.

The database contains information about patients' demographics. The demographic features contain gender (M/F), age, education, socioeconomic status and handedness. The mini-mental state examination (MMSE) is a short 30-point questionnaire test utilized to monitor for cognitive impairment and dementia. The MMSE test comprises simple questions and problems in

numeral areas: the time and place, repeating list of words, arithmetic, language utilization and comprehension, in addition with basic motor skills. Clinical dementia rating (CDR) is a numeric scale measuring the severity of symptoms of dementia. The subjects' cognitive and functional performances were accessed in six areas: memory, orientation, common sense and problem solving, community affairs, residence and hobbies, and personal care.

Table 2: Clinical dementia rating scale

CDR	Rank
0	non-dementia
0.5	very mild dementia
1	Mild dementia
2	moderate dementia

Table 3: Education codes

Code	Description
1	beneath high school graduate
2	secondary school graduate
3	some college
4	college graduate
5	above college

Table 4: Statistical data of the participants

Factor	HC	AD
No. of Patients	98	28
Age (years)	75.91± 8.98	77.75± 6.99
Education	3.26±1.31	2.57±1.31
Socioeconomic status	2.51± 1.09	2.87±1.29
CDR (0.5/1)	0	1
MMSE score	28.95±1.20	21.67±3.75
Gender(M/F)	26/72	9/19

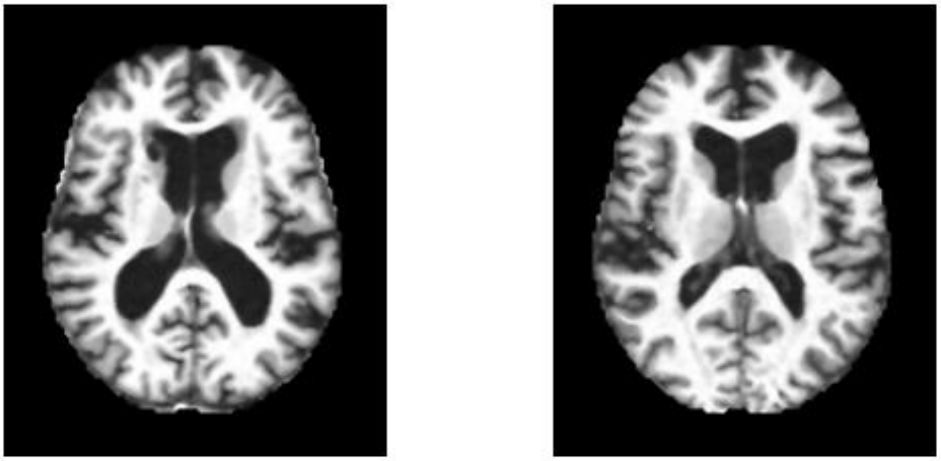


Figure10: Dataset sample (Axial view after preprocessing)

The patients' CDR ranks and education level are recorded in Table 2 and Table 3 respectively.

CHAPTER 7

Experiment, Results, and Discussion

The proposed method is implemented using the 32-bit Matlab 2015b environment on Intel(R) Core (TM) i3-2120, with a processing speed of 3.30 GHz and 2GB of RAM running Microsoft Windows 7. Readers can repeat our results on any computer with which MATLAB is compatible. This article aims at developing a CAD of AD brain system with better performance. The sample dataset is shown in Figure 12.

7.1 Parameter estimation for *S*.

It is always a major concern to find the optimum value of decomposition level *S*. We know that, a smaller S provides less information whereas a larger S provides more information to the classifier. In order to avoid overfitting problem, a smaller *S* is used. Here, I have changed the value of *S* from 1 to 5 with increment of 1, and checkup the corresponding average accuracies with feed-forward artificial neural network (FNN). The one which gives the highest accuracy is the optimal value of *S*.

7.2 Feature extraction.

In this paper, the DTCWT coefficients are extracted from the input images. The features of 5th resolution scales are selected because they provide higher classification performance than other resolution level scales. The DTCWT has a multi-resolution representation as does the wavelet transform. For disease detection, it is preferable to use a few intermediate coefficient scales as the classifier input. The lowest scales have lost fine signal details whereas the most highly detailed scales contain mostly noise. Therefore, I preferred

to choose only a few intermediate scales for the DTCWT coefficients. These obtained coefficients are sent as input to the PCA.

7.3 Feature reduction.

Excessive features increase calculation times as well as memory storage. In addition, they sometimes make classification much more complicated, which is known as curse of dimensionality. In this article, PCA is utilized to decrease the number of features. Therefore, the extracted feature from DTCWT is sent to the PCA for the feature reduction. For each image, there are 768 features after 5th level of decomposition. As, I have employed 32 slices for each patient the total number of features becomes (32×768). Now, the image is reformed into a row vector of (1×24576). The row vectors of 126 subjects are arranged into an 'input matrix' with dimension of (126×24576). It is still too large for calculation. So, the input data matrix is now decomposed into the principal component 'score matrix' and the coefficient matrix'. The score matrix size after decomposition is (126×125). Here, the rows and columns of 'score matrix' are both subjects and components, respectively.

The variance with No. of principal components from 1 to 18 are listed in Table 5. Experimenting with different numbers of principal component (PCs) revealed that accuracy with (PC=14) provided the best classification accuracy preserving 90.44% of the total variance. The curve of cumulative sum of variances with the number of principal component is shown in the Figure 11. I did not set the energy threshold as 95% because that would cost too many features, along with computational burden.

Table 5: Detailed data of PCA

No. of PCs	1	2	3	4	5	6	7	8	9
Variance (%)	63.18	72.15	77.08	80.28	83.05	84.55	85.68	86.59	87.47
No. of PC s	10	11	12	13	14	15	16	17	18
Variance (%)	88.18	88.28	89.41	89.96	90.44	90.86	91.23	91.58	91.91

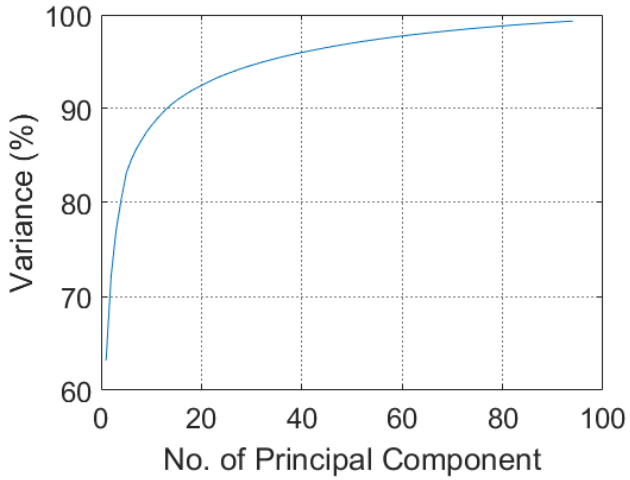


Figure 11: Variances versus No. of principal component

7.4 BPNN Training.

The 14 PC were directly sent to BPNN. Thus, the No. of input neurons (N_I) is 14. The No. of hidden layer neurons (N_H) is set to10 according to the information entropy approach [47]. Therefore, the architecture of the neural network becomes 14-10-1. The SCG method is employed because it is extremely faster than BP, MBP and ABP [12].

7.5 Statistical Analysis.

In order to execute a strict statistical analysis, stratified cross-validation (SCV) is utilized since it is a model validation approach for small-size data [48]. Moreover, SCV is also useful to make balance between reliable estimate, computational cost and for providing fair comparison. 10-SCV is employed for the dataset.

Figure 12 shows an example of K-fold SCV. The dataset used is segmented into k folds with the identical class distributions. The (K-1) folds are utilized for training, and the other remaining fold is utilized for testing, i.e., query images appears from the remaining fold. The evaluation criterion is dependent on the test images. This above technique is repeated k times so that every fold is utilized as test once.

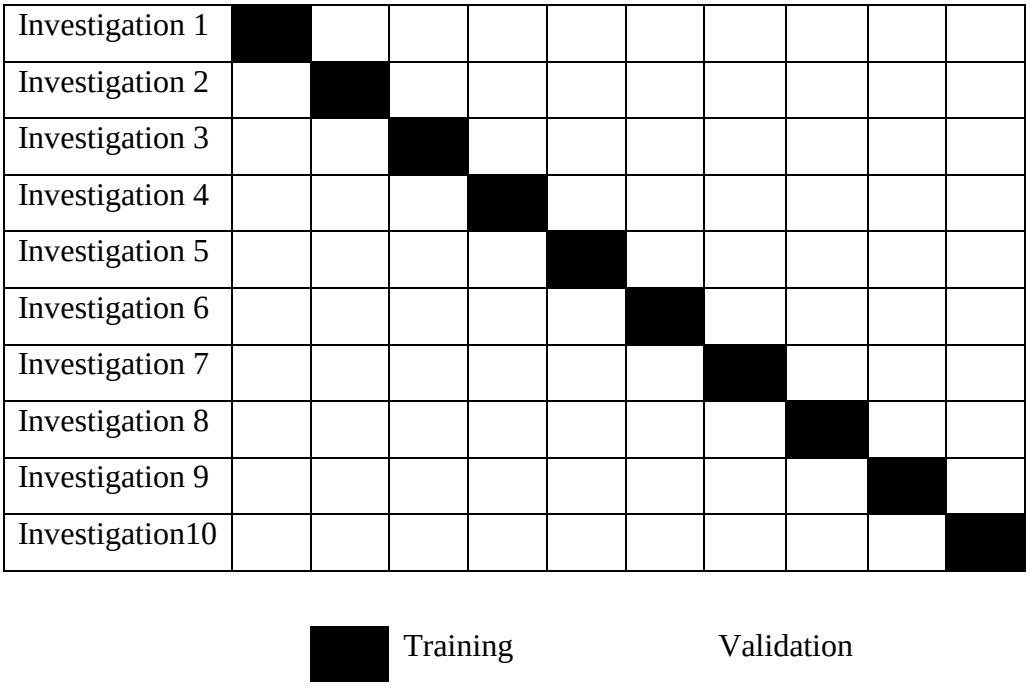


Figure 12: Illustration of 5-fold cross-validation

The final performance of K-fold SCV is acquired by averaging the K outcomes. The k-fold SCV is accomplished 10 times to further eliminate randomness.

7.6 Performance Evaluation

7.6.1 Accuracy, Sensitivity, Specificity, Precision, and Recall

There are several techniques to evaluate the efficiency of classifiers. The performance is determined on the basis of the confusion matrix. It holds the correct and incorrect classification results. The number of the instance correctly predicted by the classifier is situated on the diagonal. Table 6 demonstrates a confusion matrix for binary classification, where TP, TN, FP, and FN mean true positive, true negative, and false negative as illustrated in Table 7.

Here AD brains are assumed to hold the value “true” and NC ones are assumed to hold the value “false” following normal convention.

Table 6: Confusion matrix for a binary classifier to discriminate between two classes (A_1 and A_2).

True Class	Predicted Class	
	A_1 (Patients)	A_2 (Controls)
A_1 (Patients)	TP (True positive)	FN (false negative)
A_2 (Controls)	FP (False negatives)	TN (Ture negative)

Table 7: Evaluation indicators

Indicator	Explanation
TP	True Positive, anticipating an AD to AD
FP	False Positive, anticipating HC to AD
TN	True Negative, anticipating an HC to HC
FN	False Negative, anticipating and AD to HC

The accuracy is the most accepted empirical measure to access effectiveness of classifier. It is formulated by

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{13}$$

Sensitivity also called the true positive rate is the measure of the percentage of true positives that are correctly classified.

$$Sensitivity = \frac{TP}{TP + FN}, \tag{14}$$

Specificity also called the true negative rate is the measure of the proportion of negatives that are correctly classified.

$$Specificity = \frac{TN}{TN + FP}. \tag{15}$$

Precision is a function of true positives and instances that are misclassified as positives (false negatives).

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{16}$$

Recall is a function of its correctly classified instances (true positives) and its misclassified instances (false negatives).

$$recall = \frac{TP}{TP + FN}, \tag{17}$$

7.7. Comparison to Other state-of-the-art Approaches

To further determine the effectiveness of the proposed "DTCWT+PCA+FNN" system I have compared it with seven state-of-the-art approaches in Table 8. Some of these approaches utilized different statistical settings, making direct comparison difficult. The results in Table 8 show that study [53], [52], [28], and [54] did not present standard deviations (SD) of three standards. The specificities of study [52], [28] and [53] are lower than those demonstrated by other methods. Therefore, these three methods are not worthy of further study. Similarly, study [54] obtained a classification specificity of 100%. In spite of its high specificity, both the accuracy and sensitivity achieved by this algorithm were poor. Hence, this method is also not considered further in this study. Three other methods reported both mean values and standard deviation values. They also achieved satisfying results. Study [49] obtained promising results were achieved because of the method's voxel based morphometry (VBM). Indeed, VBM has frequently been employed to study brain changes. Study [54] demonstrated that a taxi driver will normally have a larger back section of the posterior hippocampus. Study [55] concluded that global gray matter decreases linearly with old age but global white matter remains in the same amount. Nevertheless, VBM requires an accurate spatial normalization, or the classification accuracy may decrease significantly. Study [50] was based on a novel approach called the displacement field (DF). This study measured and estimated the displacement field of various slices between AD and HC

subjects. There are other methods that have distinguished AD from HC, however, they dealt with images formed by other modalities: PET, SPECT, DTI, etc. Hence, they are also not considered in this study.

Finally, the proposed 'DTCWT+PCA+ANN' achieved an accuracy of 90.06 ± 0.01 %, a sensitivity of 92.00±0.04, specificity of 87.78 ± 0.04 %, and a precision of 89.6 ± 0.03 %. With respect to classification accuracy, our approach outperforms five other methods and is almost equal to the accuracies of the remaining two methods that did not account for means and standard deviations. A promising sensitivity and a promising specificity are also achieved. Hence, our results are either better or equivalent with other methods. The bar plot of the algorithm comparison is shown in Figure 13.

Table 8: Algorithm performance comparison for MRI brain image

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
Proposed	**90.06 ± 0.01**	**92.00 ± 0.04**	**87.78 ± 0.04**	**89.6± 0.03**
VBM+RF [49]	89.0±07	87.9±1.2	90.0±1.1	69.30±1.91
DF+PCA +SVM[50]	88.27±1.9	84.93±1.21	89.21±1.6	N/A
EB + WTT + SVM + RBF [51]	86.71±1.93	85.71±1.91	86.99±2.3	66.12±4.16
BRC+IG+SVM [52]	90.00	96.88	77.78	N/A
BRC+IG+VFI[52]	78.00	65.63	100.00	N/A
Curvelet + PCA + KNN [53]	89.47	94.12	84.09	N/A
US+SVD-PCA+SVM-DT [54]	90.00	94.00	71.00	N/A

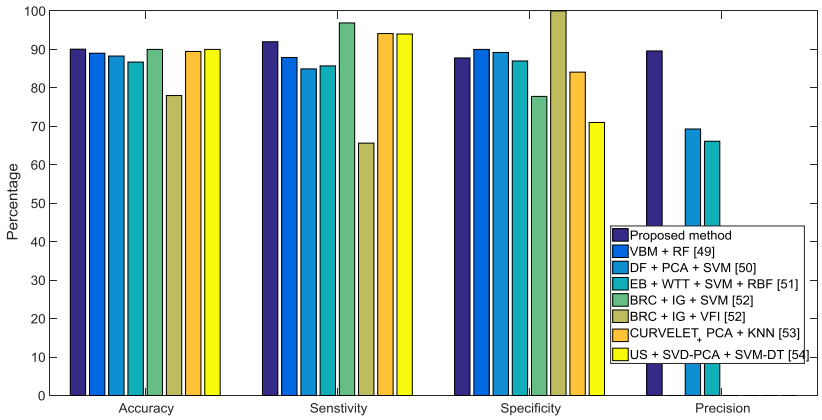


Figure 13: Bar plot of the proposed system.

7.8 Computation time

Computational time is another major factor to analyze the classifier. The network training time are not taken into consideration because the weights/biases of FNN should keep unchanged unless the image property alters much. I have sent all the 126 subjects into classifier and noted the corresponding computational time, calculated the mean value, and represent consumed time of various stages as shown in the Figure 14.

For each image, the computational time for feature extraction, feature reduction, and feed-forward based classifier is 178.20s, 72.48 s, and 13.05s, respectively. It is seen form the experiment that feature extraction stage consumes more time as compared to others, that must be enhanced in future research. The total computational time for each image is about 15 sec, which is rapid enough the real time diagnosis.

Table 9: Computational Time

Process	Time (second)
DTCWT	178.20
PCA	72.48
FNN	13.05

Table 9 shows the computational time for the each of the steps. The total time is 263.73s .Therefore, the proposed system in feasible in practice.

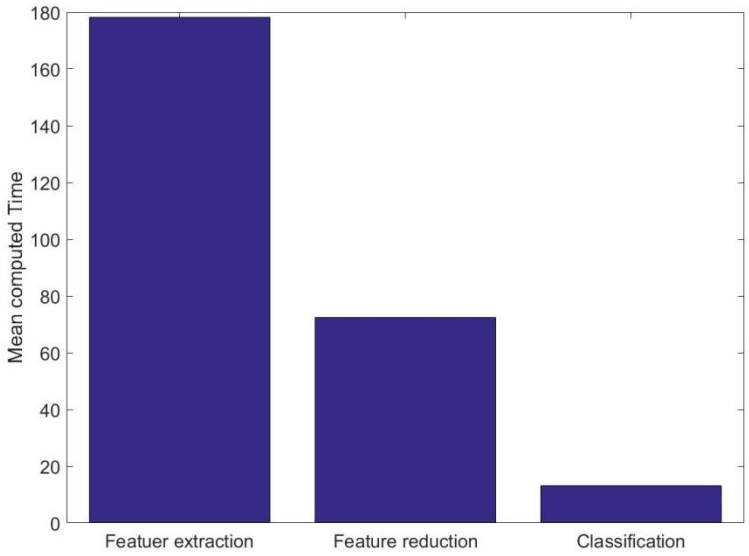


Figure 14: Offline-Learning Computation Time

8. CONCLUSIONS AND FUTURE RESEARCH

I presented an automated and accurate method for AD identification based on a dual tree complex wavelet transform (DTCWT), principal component analysis (PCA), and a feed-forward neural network. The results showed that the proposed method achieved an accuracy of 90.06 ± 0.01%, a sensitivity of 92.00 ± 0.04 %, and a specificity of 87.78 ± 0.04 % and precision of 89.6 ± 0.03 % and outperformed 7-state-of-the-art-algorithms.

our future research shall focus on the following aspects: (i) testing other advanced variants of wavelet such as 3D-DTCWT, wavelet packet analysis and fractional calculus, (ii) utilizing different feature reduction techniques such as autoencoders, linear discriminant analysis (LDA), and Probabilistic PCA,(iii) testing our data with least-square techniques, kernel support vector machine (SVM), and deep learning methods such as convolutional neural network (CNN) and other alternative pattern recognition tool for classification, (iv) utilizing advanced swarm intelligence techniques such as artificial bee colony, particle swarm optimization (PSO), genetic pattern search, ant colony optimization, bibliography based optimization (BBO) to find the optimal kernel. (v) testing the proposed method on images obtained from different modalities such as computed tomography ultrasound, spectrum imaging,3-dimensional MRI etc. (vi) utilizing other advance image preprocessing technique to enhance the classification performance, such as image de-noising, image enhancement, and image segmentation. (vii) Classification may be carried out on the sparsity domain.

REFERENCES

[1] Y. Zhang, Z. Dong, A. Liu, and S. Wang, G. Ji, Z. Zhang, and J. Yang, "Magnetic resonance brain image classification via stationary wavelet transforms and generalized eigenvalue proximal support vector machine," Journal of Medical imaging and Health Informatics, vol. 5, no. 7, pp. 1395-1403, 2015.

[2] S. Goh, Z. Dong, Y. Zhang, S. DiMauro, and B. S. Peterson "Mitochondrial dysfunction as a Neurobiological Subtype of Autism Spectrum Disorder: Evidence From Brain Imaging," JAMA Psychiatry, vol.71, no. 6, pp. 665-671, 2014.

[3] 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.

[4] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski, "Prediction of MCI to AD conversion, via MRI,CSF biomarkers, and pattern classification," Neurobiology of Aging, vol. 32, pp. 2322.e19–2322.e27, Dec. 2011.

[5] A. Nordberg, J. O. Rinne, A. Kadir, and B. Langstrom, "The use of PET in Alzheimer disease," Nature Revolution, Neurology, vol.6, no. 2, pp. 78–87, Feb. 2010.

[6] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI," Proc. Nat. Acad. Sci. United States Amer., vol. 101, no. 13, pp. 4637–4642, Feb. 4, 2004.

[7] T. Magnander, E. Wikberg, J. Svensson et al., "A novel statistical analysis method to improve the detection of hepatic foci of 111 In-octreotide in SPECT/ CT imaging," EJNMMI Physics, vol. 3, no.1, pp. 1, 2016.

[8] P. D. Visschere, M. Nezzo, E. Pattyn, V. Fonteyne, C. V. Praet and G. Villeirs "Prostate magnetic resonance spectroscopic imaging at 1.5 tesla with endorectal coil versus 3.0 tesla without endorectal coil: comparison of spectral quality," Clinical Imaging, vol. 39, pp. 636-641, 2015.

[9] 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.

[10] Z. Dong, Y. Zhang, F. Liu, Y. Duan, A. Kangarlu, B. S. Peterson, "Improving the spectral resolution and spectral fitting of 1H MRSI data from human calf muscle by the spread technique," NMR in Biomedicine, vol. 27, no. 11, pp. 1325-1332, 2014.

[11] C. Westbrook, Handbook of MRI technique, John Wilety & Sons, oxford, 2014.

[12]Y. Zhang, Z. Dong, L. Wu, and S. Wang, "A hybrid method for MRI brain image classification," Expert Systems with Applications, vol. 38, no. 8, Aug. 2011.

[13]S. Chaplot, L. M. Patnaik, 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.

[14]M. Maitra, and A. Chatterjee, "A Slantlet transform based intelligent system for magnetic resonance brain image classification, Biomedical Signal Processing and Control, vol. 1, no. 4 , pp. 299-343, 2006.

[15]E. S. A, El-Dahshan, T. Hosny, A. B. M, Salem, "Hybrid intelligent techniques for MRI brain images classification," Digital Signal Processing, vol. 20, no. 2, pp. 433-441, Mar. 2010.

[16] C. Plant, S. J. Teipel et al., "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.

[17]Y. Zhang, Z. Dong, L. Wu, and S. Wang, "A hybrid method for MRI brain image classification," Expert Systems with Applications , vol. 38, no. 8, pp. 10049-10053, 2011.

[18] R. Ramasamy, and P. Anandhakumar, "Brain Tissue Classification of MR Images Using Fast Fourier Transform Based Expectation-Maximization Gaussian Mixture Model," Advances in Computing and Information Technology, pp. 387-398, 2011.

[19] M. Sairtha, K. P. Joseph, and A. T. Mathew, "Classification of MRI brain images using combined wavelet entrophy based spider plots

and probabilistic neural network," Pattern Recognition Letters, vol. 34, no. 16, pp. 2151-2156, 2013.

[20] A. Savio, M. Grana, "Deformation based feature selection for computer aided diagnosis of Alzheimer's disease," Expert Systems with Applications, vol. 40, no. 5, pp. 1619-1628, 2013.

[21] S. Das, M. Chowdhary, and M. K. Kundu,"Brain MR Image Classification Using Multiscale Geometric Analysis of Ripplet," vol. 137, pp. 1-17, 2013.

[22] H. Kalbkhani, M. G. Shayesteh, B. Z. Vargahan, "Robust algorithm for brain magnetic resonance image (MRI) classification based on GARCH variances series," Biomedical Signal Processing and Control, vol. 8, no. 6, pp. 909-919, 2013.

[23] Y. Zhang, S. Wang, and Z. Dong, "Classification of Alzheimer' Disease based on Structural Magnetic Resonance Imaging by Kernel Support Vector Machine Decision Tree," vol. 144, pp. 171-184, 2014.

[24] El-Sayad, El-Dashan et al., " Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm," Expert Systems with Applications, vol. 41, no. 22, pp. 5526-5545, 2014.

[25] [25] X. Zhou, S. Wang, W. Xu, et al., "Detection of Pathological Brain in MRI Scanning Based on Wavelet- Entropy and Naïve Bayes Classifier," International Conference on Bioinformatics and Biomedical Engineering, vol. 9043, pp. 201-209, 2015.

[26] Y. Zhang, Z. Dong, S. Wang, G. Ji and J. Yang, " Preclinical Diagnosis of Magnetic Resonance (MR) Brain Images via Discrete Wavelet Packet Transform with Tsallis Entropy and Generalized Eigenvalue Proximal Support Vector Machine (GEPSVM)," Entropy, vol. 14, no. 4, pp. 1795-1813, 2015.

[27] G. Yang, Y. Zhang, J. Yang, G. Ji, Z. Dong, S. Wang et al., "Automated classification of brain images using wavelet-energy and biogeography-based optimization," Multimedia Tools and Applications, vol. 75, no. 23, pp. 15601-15617, 2016.

[28] D. Jha and G-R kwon, "Ensembles of Decision Trees Based Diagnosis of Alzheimer's Disease and Dementia," The Journal of Korean Institute of Next Generation Computing, vol. 13, no. 2, Apr. 2017.

[29] D. Jha and G-R Kwon, “Alzheimer’s disease detection using sparse autoencoder, scale conjugate Gradient and softmax output layer with fine tuning,” International conference on Frontiers of Image processing, March 2017.

[30] Y. Zhang, Z. Dong, L. Wu and S. Wang, “A hybrid method for MRI brain image classification,” Expert Systems with Applications, vol. 38, no. 8, pp. 10049-10053, 2011.

[31] P. S. Hiremath, S. Shivashankar, and J. Pujari, “Wavelet based features for features for color texture classification with application to CBIR, International Journal of Computer Science and Network Security,” vol. 6, no. 9(A), pp. 124-133, 2006.

[32] D. Bouchaffra, and J. Tan, “Structural hidden Markov models for biometrics: Fusion of face and fingerprint,” Pattern Recognition, vol. 41, no. 3, pp. 852-867, Mar. 2008.

[33] K. Roy, and P. Bhattacharya, “optimal features subset selection and classification for Iris recognition, journal on Image and Video Processing, 2008.

[34] N. Kingsbury, “Complex wavelets for shift invariant analysis and filtering of signals,” Applied and Computational Harmonic Analysis, vol. 10, no. 3, pp. 234-353, 2001.

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

[36]S. Wang, S. Lu, Z. Dong et al., “Dual-Tree Wavelet Transform and Twin Support Vector Machine for Pathological Brain Detection,” Applied Science, vol. 6, no. 6, pp. 169, 2016.

[37]A. K. Jain, R. P. W. Dulin, and J. Mao, “Statistical pattern recognition: A review,” IEEE Trans. Pattern Anal. Mach. Intell. vol. 22, pp. 4-37, 2000.

[38] C. M. Bishop, “Pattern recognition and machine learning,” Springer, New York.

[39] R. O Duda, P. E. Hart, and D. G. Stork, “pattern classification,” Wiley, New York, 1973.

[40]M. E. Tipping, and C. M. Bishop, “Probabilistic principal component analysis’, Journal of the Royal statistical Society: Series B” Statistical Methodology, vol. 61, no. 3, pp. 611-622, 1999.

[41] S. Roweis, “EM Algorithms for PCA and SPCA,” Advance in neural information proccessing System, 10, (MIT Press, Cambridge, MA, USA, 1998), pp. 626–632, 1998.

[42] A. Ilin, and T. Raiko, “Practical Approaches to Principal Component Analysis in the Presence of Missing Values,” Journal of Machine Learning Resource, vol. 11, pp. 1957–2000, 2010.

[43]D. L. Guo, Y. D. Zhang, and Z. Li, “Improved Radio Frequency Identification Method via Radial Basis Function Neural Network,” Mathematical Problems in Engineering, 2014.

[44]M. Manoochehri and F. kolahan “Integration of artificial neural network and simulated annealing algorithm to otimize deep drawing process,” The International Journal of Advaned Manufacturing Technology, vol. 73 , no. 1, pp. 241-249, 2014.

[45]S. U. Aswathy, G. G. Deva Dhas and S. S. Kumar, “A survey on Detection of Brain Tumor from MRI Brain Images,” contol, Instrumentation, Communication and Computational Technologies (ICCICCT), International conference, 2014.

[46]A. Poursamad,“Adaptive feedback linearization control of anticlock braking systems using neural networks,” Mechatronics, vol. 19, no. 5, pp. 767-773, 2009.

[47]H. C. Yuan, F. L. Xiong, and X. Y. Huai, “A method for estimating the number of hidden neurons in feed-forward neural networks based on information entropy,” Computers and Electronics Agriculture, vol. 40, no. 1-3, pp. 57-64, 2003.

[48]S. Purushotham, and B.K Tripathy, “Evaluation of Classifier Models Using Stratified Tenfold Cross Validation Techniques,” Global Trends in Information Systems and Software Applications, pp. 680-690, 2012.

[49]K. R. Gray, P. Alijabar, R. A. Heckemann, A. Hammers, D. Rueckert, “Random forest-based similarity measures for multi-modal classification of Alzheimer’s disease,” NeuroImage, vol. 65, pp. 167–175, 2013.

[50] Y. Zhang and S. wang, "Detection of Alzheimer's disease by displacement field and machine learning," PeerJ, vol. 3, 2015.

[51] Y. Zhang, Z. Dong, P. Phillips, S. Wang, Ji. Genlin, J. Yang, T-F Yuan, "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," Frontlier in Computational Neuroscience, vol. 9, 2015.

[52] C. Plant, S. J. Teipel, A. Oswald, C. Bohm, T. Meindl, J. M. Miranda, A.W. Bokde, H. Hampe, and M. Ewers, "Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease," NeuroImage, vol. 50, pp. 162–174, 2010.

[53] D. Jha and G-R kwon, "Alzheimer's Disease detection in MRI Using Curvelet Transform with K-NN," Journal of KIIT, vol. 14, no. 8, Aug. 31, 2016.

[54] Y. Zhang, S. Wang, and Z. Dong, "Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree," Progress In Electromagnetic Research, vol. 144, pp. 171–184, 2014.

[55] E. A. Maguire, D. G. Gadian, I. S. Johnsrude, C. D. Good, J. Ashburner, R. S. J. Frackowiak, and C. D. Frith, "Navigation-related structural change in the hippocampi of taxi drivers" Proceeding of the National Academy of Sciences, vol. 97, no. 8, pp. 4398-4403, 2000.

[56] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. A. Henson, K. J. Friston, and R. S. J. Frackowiak, "A voxel-based morphometric study of ageing in 465 normal adult human brains," Biomedical Imaging, 2002.

LIST OF PUBLICATIONS

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.

Debesh Jha and Goo-Rak Kwon, "Alzheimer's Disease Detection in MRI Using Curvelet Transform with K-NN," The Journal of Korean institute of Information Technology, Vol. 14, No. 8, Aug. 2016.

Debesh Jha and Goo-Rak Kwon,"Classification and Diagnostic prediction of Alzheimer's and different types of dementia," KSBNS intelligent Brain, Sept. 2016.

Debesh Jha and Goo-Rak Kwon, "Alzheimer's disease detection using sparse autoencoder, scale conjugate gradient and softmax output layer with fine tuning," International Conference on Frontiers of Image Processing, Mar. 2017.

Debesh Jha and Goo-Rak Kwon, "Diagnosis of Alzheimer's Disease Using Dual-tree Complex Wavelet Transform Wavelet Transform, PCA, and Feed-forward Neural network," Journal of Healthcare Engineering, May. 2017. (In press)

Debesh Jha and Goo-Rak Kwon, "Ensembles of Decision Trees based diagnosis of Alzheimer's disease and Dementia," Journal of Korea Convergence Society, May, 2017.

Debesh Jha and Goo-Rak Kwon,"A hybrid model based for pathological brain image detection from MRI," Journal of Medical Imaging and Health Informatics, 2017 (Major Revision).

Debesh Jha and Goo-Rak Kwon,"Brain image segmentation method based on dual-tree complex wavelet Transform and fuzzy c-means clustering algorithm," Journal of healthcare engineering, 2017. (Under review)

Debesh Jha and Goo-Rak Kwon, "Pathological Brain Detection Using Weiner filtering, 2D- Discrete wavelet Transform, Probabilistic PCA and Random Subspace Ensemble Classifier," Computational Intelligence and Neuroscience, 2017 (**Under review**).