Islamic University – Gaza

Deanery of Higher Studies

Faculty of Engineering

Computer Engineering Department

Brain Tumor Detection Based On Curvelet and Artificial Neural Network

Muhammad R. Al kahlout

Supervisor:

Prof. Mohammad Mikki

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Computer Engineering

1435H (2013)

ACKNOWLEDGMENT

My thanks go to all people that were involved in my master's thesis. I would like to express my immense gratitude to my supervisor, Professor Mohammed Mikki, for his advice, support, guidance and inspiration to conduct my research.

I gratefully acknowledge and express my deep thanks to the discussion committee members, Professor Samy Abu Naser and associate professor Mohammed Alhanjouri, for their valuable comments, suggestions and recommendations.

I would also like to thank all my friends and my colleagues in the computer engineering department for their friendship and assistance during my research study. Special thanks t for their assistance and expert advice.

My heartiest gratitude to my beloved family: my wife, mother, father and children, without their support, understanding, and patience, it would have been impossible for me to finally complete my study.

Above all, I thank Allah for blessing me with all these resources, favors and enabling me to complete this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANFIS   :Adaptive Neuro Fuzzy Inference System

ANN   :Artificial Neural Network

BPN   :Back Propagation Neural Network

CADx   :Computer Aided Diagnosis

CFD   :Curvelet Feature Descriptor

CSF   : Curvelet Statistical Features

CT   :Computed Tomography

DWT   :Discrete Wavelet Transform

EM   :Expectation Maximization

FCM   :Fuzzy C-means

FF-ANN   :Feed Forward Artificial Neural Network

FFT   :Fast Fourier Transform

GA   :Genetic Algorithm

GLCM   :Grey Level Co-occurrence Matrix

IFFT   :Inverse Fast Fourier Transform

KNN   :K-Nearest Neighbors

LDA   :Linear Discriminant Analysis

MRA   :Multi Resolution Analysis

MRI   :Magnetic Resonance Imaging

MSE   :Minimum Square Error

PCA   :Principal Component Analysis

PDP   :Parallel Distributed Processing

PNN   :Probabilistic Neural Network

PE   : Processing Element

PET   :Positron Emission Tomography

RBFN   :Radial Basis Function Neural Network

SVM   :Support Vector Machines

SWT   :Stationary Wavelet Transform

USFFT   :Unequally Spaced Fast Fourier Transform

WHO   :World Health Organization

نظام تشخيص أورام الدماغ باستخدام الشبكات الاصطناعية ومحول مصغر المنحنى

محمد راجح الكحلوت

ملخص

إن الكشف عن الأورام في الصور الطبية المختلفة بإستخدام نظم التصنيف الطبي الآلي يتطلب دقة عالية لأنه يتناول حياة الإنسان. يعتبر التصوير بالرنين المغناطيسي (MRI) من أفضل طرق التصوير الطبية في أنظمة التشخيص الطبي الآلي، وذلك لأنه يوفر قدرا كبيراً من التباين بين الأنسجة الرخوة المختلفة في جسم الإنسان وبالتالي التصوير بالرنين المغناطيسي هو أكثر فعالية في تشخيص أورام الدماغ. يعتمد الأطباء وأخصائيي الأشعة على أنظمة التشخيص الطبي الآلي (CADx) لمساعدتهم في الكشف عن الأورام في صور الرنين المغناطيسي للدماغ حيث تكون بمثابة رأي مساند أومرجح. لذلك فإن التحسين في أداء (CADx) يزيد من خيارات الكشف المبكر والعلاج والتماثل للشفاء.

إن الهدف الرئيسي من بحثنا هو تحسين وإدخال طريقة جديدة لإستخراج ميزات وخصائص الصور من أجل بناء نظام تشخيص طبي (CADx) يعطي نتائج موثوقة في تصنيف صور الرنين المغناطيسي للدماغ. لقد وضعنا تركيزنا في هذه الأطروحة على استخراج ميزات النسيج، ميزات الملمس هي واحدة من الخصائص الأكثر أهمية للصورة، فهي تميز نمط سطح الأجسام في الصور. ميزات الملمس تصف بشكل فعال الخصائص المميزة بين الصور.

بعد دراسة مستفيضة لمميزات وعيوب طرق استخراج المميزات من الصور وجدنا أن تحويل مصغر المنحنى (Curvelet) هو أفضل طريقة لتمثيل حافة المنحنيات في الصور الطبية وأقلها تأثراً بالتشويه الموجود في الصور.

في أطروحة الماجستير هذه نقدم نظاماً لتشخيص أورام الدماغ باستخدام الحاسوب وصور الرنين المغناطيسي و يتكون من ثلاثة مراحل ، خلال المرحلة الاولى يتم استخراج ميزات النسيج لوصف محتوى الصورة بإستخدام واحد من أحدث تقنيات التحليل متعدد النتائج لإستخراج ميزات صورة الرنين المغناطيسي للدماغ و يدعى تحويل مصغر-المنحنى (Curvelet) في المرحلة الثانية نستخدم تقنية تحليل المكونات الرئيسية (PCA). وذلك لتقليل عدد الخواص الممثلة للصورة واختيار أفضلها مما يساعد على زيادة سرعة ودقة أداء النظام ، في المرحلة الأخيرة يتم إدخال ناتج الخواص الممثلة للصورة الي أكثر طرق التعرف على الانماط إستخداماً وهو الشبكة العصبية الإصطناعية (ANN) لتصنيف الصورة.

لقد أجرينا التجارب في هذا البحث على مجموعة من قواعد البيانات المعروفة عالمياً في هذا المجال، وتتكون هذه المجموعة من 160 صورة رنين مغناطيسي للدماغ مصنفة بمعرفة أطباء في كلية الطب جامعة هارفارد إلي 132 حالة ورم خبيث و 28 حالة سليمة. وقد لاحظنا أن النظام باستخدام المميزات التي حصلنا عليها باستخدام تحويل مصغر المنحنى (curvelet) قد أعطى نتائج تنصيف كالتالي % 96.4 من حالات الأورام الخبيثة و %100 من الحالات السليمة قد تم تصنيفها بشكل صحيح ونسبة الدقة وصلت الى %96.9.

ABSTRACT

An Automated classification and detection of tumors in different medical images demands high accuracy since it deals with human life. In automated medical diagnostic systems, magnetic resonance imaging (MRI)gives better results than any other imaging modalities as it provides greater contrast between different soft tissues of human body. Hence MRI is much more effective in brain cancer imaging

Computer-aided diagnosis (CADx) is used to help physicians and radiologists in classification of MRI brain image and is usually used as a second opinion by the radiologists.

Improving CADx increases the treatment options and a cure is more likely. The main objective of our research is to enhance and introduce a new method for feature extraction and selection in order to build a CADx system to discriminate between normal and abnormal cases.

Thus we put our focus in feature extraction, texture features is one of the most important and prominent properties of an image. It is the surface pattern of objects in the image or the whole image. Texture features effectively describe the distinguishing characteristics between images.

After extensively studying the advantages and disadvantages of several spectral methods of texture feature descriptors such as wavelet transform and Gabor filter we have adopted a wrapping based discrete curvelet texture descriptor for use in CADx. Discrete curvelet transform is one of the most powerful approaches in capturing edge curves in medical image

In this research, we propose a new model for automatic brain tumor diagnosis system from MR images. The system consists of three stages namely feature extraction ,feature selection and classification . In the first stage, we have obtained the curvelet texture features descriptor (CFD) related to MRI images using Fast Discrete Curvelet transformation (FDCT). In the second stage, the features vector of brain MRI have been reduced to the more essential features using principal component analysis (PCA), eventually classification stage, We use feed forward artificial neural network with backpropagation (FP-ANN).

We generate a texture features descriptor using wrapping based discrete curvelet transform. This descriptor is used to represent images database. The optimal level of curvelet decomposition is also investigated to obtain the highest recognition rate, CADx tests are performed using database containing 160 images obtained from Faculty of medicine in Harvard university. Experimental results for the proposed method achieved an overall classification accuracy of 96.9%, with 96.4% sensitivity and 100% specificity.

***Key words*:** Brain Tumor , Curvelet , Feature Extraction, Feature Selection, Computer Aided Diagnosis, Principal Component Analysis , Artificial neural   network.

Chapter 1
Introduction

1.1 Background

Cancer is the second leading cause of death for both men and women in Palestine and is expected to become the leading cause of death in the next several decades[1]. It has been shown that early detection and treatment of brain cancer are the most effective methods of reducing mortality.

The rapid development in image processing and soft computing technologies have greatly enhanced interpretation of medical images, and contributed to early diagnosis.

Therefore, how to detect, analyze, and treat cancers has become a big research field. Modern imaging technology as shown in Fig. 1.1 has already had lifesaving effects on our ability to detect cancer early and more accurately diagnose the disease. However, in order to further improve the efficiency and accuracy of diagnoses and treatment, image processing technology has been widely applied to analysis and recognition of cancer, evaluation of the treatment effectiveness, and prediction of the development of cancer.
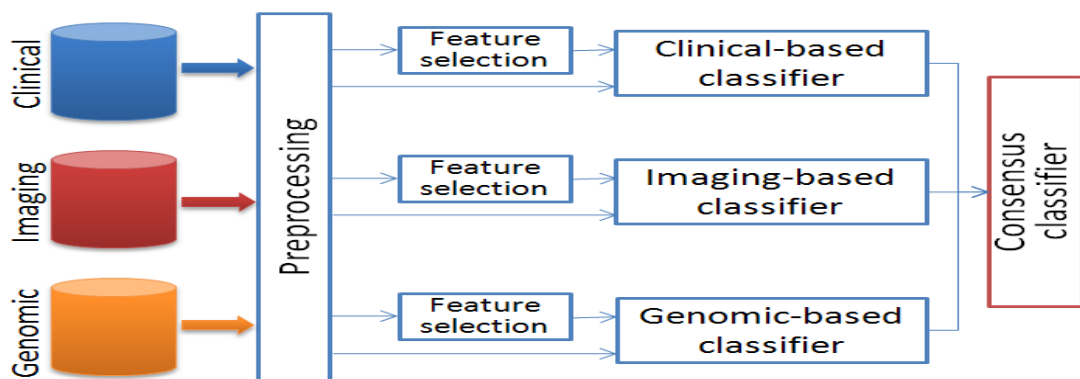
Figure 1.1:Types of Classification.

A Brain Tumor is uncontrolled growth of tissues within the brain; any brain tumor is inherently serious and life-threatening disease [2]. Brain tumor is classified based on malignancy by the World Health Organization (WHO). There are over 120 types of

brain and central nervous system tumors. They can be malignant or non-malignant (benign), and in either case, can be just as injurious or life threatening. Tumors are classified based on the cell origin and behavior of cells. A brain tumor which has started in the brain is a primary tumor. Tumors that metastasize into the brain are secondary brain tumors. Most physicians use the WHO's classification system to identify type of brain tumor. The most common types of primary brain tumors are: Glioma, Meningioma, Pituitary adenoma and Nerve sheath tumors. Classification of brain tumors based on their malignancy scale is as follows in Fig 1.2:

Grade (I) at the first row of Fig 1.2 tumor cells appear normal and are benign. The tumor is slow growing. Grade (II) at the second row of Fig 1.2 tumor is malignant. The cells show changes from their normal appearance. At the third row of Fig 1.2 Grade (III) tumor cells are very different from normal brain cells and tumor is malignant and fast growing. Finally Grade (IV) tumor cells are abnormal in appearance. Malignant tumor is fast growing.
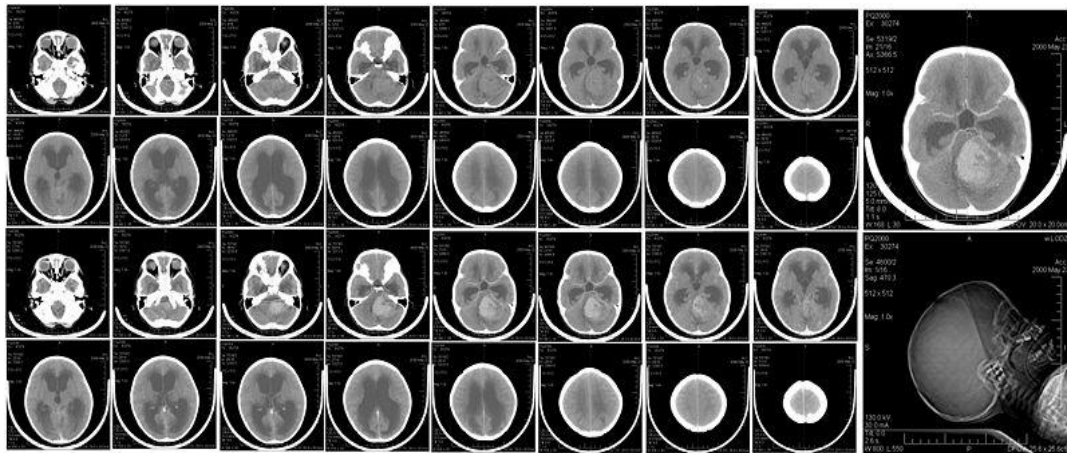
Figure 1.2: Classification of brain tumors based on their malignancy.

Over time, a slow growing low grade tumor can become a fast growing malignant tumor. Therefore the early detection and classification of brain tumor based on its grade and malignancy help predict prognosis and course of treatment.

Medical Imaging plays a central role in the diagnosis of brain tumors[3]. Magnetic Resonance Imaging (MRI) is one of the best technologies currently being used for diagnosing brain tumor. MRI imaging uses a powerful magnetic field, radio

frequency pulses and a computer to produce detailed pictures of organs and soft tissues. It helps physicians to treat medical conditions. MRI is preferred because it is more comfortable than computed tomography (CT) scan for diagnosis. Moreover it has no side effects on the human body because it doesn't use any radiation that it is based on the magnetic field and radio waves. Also it gives better results than (CT) as it provides greater contrast between different soft tissues of human body [3].

Medical images (including MRI) are strongly deteriorated by noise mainly due to data acquisition systems, various sources of interference, operator performance and other phenomena that affect the measurement processes in imaging which can lead to serious classification inaccuracies. Visual interpretation of brain tumor by doctors and radiologists is a tedious and fatiguing process especially when a large number of patients in intensive care units need for continuous observation. This process generally requires a magnifying glass and perhaps leads to different and inaccurate diagnosis, but when dealing with human life this is unacceptable. So the necessity of high accuracy medical diagnosis system or computer aided diagnosis system (CADx) is highlighted. CADx system is demanded in medical institutions due to the fact that it could support the decision of radiologists and doctors by providing a computer result as a second opinion.

Thus designing a completely automatic and efficient medical diagnosis system is a grand challenge. Generally such a medical diagnosis system is able to perform three main sequentially subtasks namely feature extraction, feature selection and classification.

- **Feature extraction**: this step is responsible for extracting all possible features in medical image MRI that are expected to be effective in diagnosing brain tumor , without concerning the disadvantages of excessive dimensionality.

- **Feature selection**: this step is responsible for reducing the dimensionality by eliminating irrelevant or redundant features and searching for the best significant features to avoid the curse of dimensionality and reducing computation complexity.

- **Classification:** problem occurs when an image needs to be assigned into a predefined group or class based on a number of observed features. Fig. 1.3 shows the typical architecture of the CADx system.

Figure 1.3: Typical architecture of a CADx system.

1.2 Research Overview

In this section, we present detailed information about this thesis. We start by identifying the importance of the optimization of the feature extraction for MRI brain tumor diagnosis including the motivations behind our study, objectives to be accomplished, methodology that has been followed, our contributions throughout this work, and finally, we show the content of this research.

1.2.1 Motivation

As mentioned before, early detection and diagnosis of brain tumor increase the treatment options and a cure is more likely. CADx is an active tool used for early detection but 10%-30% of patients who have the disease and undergo diagnosis have negative classification[4]. Two thirds of these false negative cases were evident retrospectively. These mistakes in the visual interpretation are due to poor image quality, eye fatigue of the radiologist, subtle nature of the findings, or lack experienced radiologists especially in third-world regions. Nowadays the computer-aid systems play the main role in early detection and diagnosis of brain tumor. Increasing confidence in the diagnosis based on CADx systems would, in turn decrease the number of patients with suspected brain cancer who have to undergo surgical brain biopsy, with its associated complications.

1.2.2 Objectives

The performance of CADx system depends more on the optimization of the feature extraction method than the classification methods [5]. Our objective in this thesis is to design, develop and evaluate an easy to use automated intelligent medical diagnosis system. See the impact of using Curvelet as feature extraction tool, exploiting the characteristics of Curvelet for image denoising, scale distortion and solving curve singularities. However the feature space is very large and complex due to the wide diversity of the normal tissues and the variety of the abnormalities. Using excessive features may degrade the performance of the algorithm and increase the complexity of the classifier. For this, the main goal of this research is to evaluate methods for feature selection Principal Component Analysis (PCA) that guarantee the enhancement of classification.

1.2.3 Contributions

The Contributions of this research is list as follows:

- We propose a new hybrid scheme for (MRI) brain image Classification.
- We design and develop an effective and efficient computer aided diagnosis system for brain tumor diagnosis using Fast discrete curvelet transform (FDCT) as feature extraction tool, Principal component analysis (PCA) as dimensionality reduction tool and feed forward artificial neural network (FF-ANN) for classification.
- We use Curvelet transform, which is a powerful texture extraction technique in describing image content at different resolutions. we generate a texture features descriptor that can be used to represent images in a large database in terms of their features.
- We enhance the feature extraction stage via utilizing the special characteristics of the curvelet transform for noise removal during the process of feature extraction, a higher curvelet coefficient corresponds to a stronger edge, and the noise corresponds to a smaller coefficient. So when choosing the proper threshold, we may retain the bigger coefficient and abandon the smaller coefficient to achieve the image denoising.

- Furthermore, the optimal level of curvelet decomposition is also investigated to obtain the highest outcome in terms of sensitivity, specificity and accuracy.

- We have improved the efficiency of CADx system thus, by reducing computational cost of curvelet transform by applying a method for optimizing number of scales and wedges depending on the image size.

- We use the principal component analysis (PCA) to Select image features that will efficiently represent the image and greatly reduce the computation complexity.

- We improve medical CADx system with sufficient generalization of performance and high classification accuracy, with adequate system speed and low computation load.

1.2.4 Organization of This Thesis

The rest of the thesis is organized as follows. Chapter 2 summarizes some of the related works in the topic of feature extraction, feature reduction and classification techniques for computer aided diagnosis system. In Chapter 3, we introduce an overview of the CADx systems, its principle, and the theoretical foundations of the techniques used for feature extraction, feature reduction and classification. Fast Discrete Curvelet transform , the essential technique used for feature extraction in our proposed systems, is also discussed in this chapter. In Chapter 4, we introduce our proposed system for CADx based on curvelet, we describe the method to compute a curvelet texture descriptor for image databases. In addition, the dimension reduction of feature space process based on principal component analysis technique and finally the classification process using artificial neural network Chapter 5 analyzes the effectiveness and efficiency of curvelet texture features on both normal and scaled image classification. The outcome is compared with two well established texture representations, i.e., discrete wavelet and Gabor filters. The system performances of discrete curvelet, discrete wavelet and Gabor filters features in CADx are analyzed in this chapter.

Finally, Chapter 6 concludes the thesis with a summary of the findings and contributions in this research. Several possible future directions of this research are also provided in this chapter.

Chapter 2
Related Work

2.1 Introduction

Developing an effective computer-aided diagnosis (CADx) system for Brain tumor detection is of great clinical importance and can increase the patient's chance of survival. For this reason, CADx systems for brain tumor have been investigated in a huge number of research studies [6].

 A typical CADx system for brain cancer diagnosis is composed of four main processing steps namely Preprocessing of the Medical image, Feature Extraction of medical image, Feature Reduction to reduce the search space and Classification of the medical image as normal or abnormal

With the advances in soft computing and image processing. universities and other research centers all around the world, started doing research on the field of computer aided medical image diagnosis.

 The use of medical image processing has been increased especially for medical applications, this leads to increase the qualified radiologists number who navigate, view, analyze, segment, and interpret medical images [7]. The analysis and visualization of the medical image received from the acquisition devices such as PET, CT, or MRI are difficult to evaluate due to the quantity of clinical data and the amount of noise existing in medical images due to the scanners itself. Computerized analysis and automated information systems can offer help dealing with the large amounts of data, and new image processing techniques may help to denoise those images. CADx depends on several factors, such as, feature extraction method, dimension reduction method that produce the most discriminative features and Classification method. Designing an efficient CADx system for Medical image diagnosis is still challenging. Important factors should be investigated including the automation level, the speed, the ability to detect tumors of different shapes, for example, irregularly shape tumors not only the spherical ones, and the ability of the CADx system to detect cavity and small tumors that attached to the organ borders.

Since an improvement to any of these influencing factors can result in a more effective diagnosis system. For this purpose, we first provide a brief review of the factors that can affect CADx.

2.2 Feature Extraction and Dimension Reduction

Feature extraction of medical images is a crucial step in any CADx system that can lead to the early diagnosis of abnormalities [9]. Texture analysis is one of the most important techniques used in the analysis and interpretation of images. Texture feature analysis methods can be classified into four primary categories, namely statistical, geometrical, model-based and signal processing-based approaches.
The signal processing-based approach has been widely used in the field of texture feature extraction [8]. Which allows for the preservation of an image according to certain levels of resolution Broadly speaking, multi-resolution analysis (MRA) allows for the zooming in and out of the underlying texture structure. In texture analysis, the most difficult aspect is to define a set of meaningful features. Recently, one of the major developments in texture classification has been the use of multiresolution analysis. Multiresolution analysis provides information about the image contained in both time and frequency domain, and thus provides a powerful tool for the description of similar textures. Several multiresolution transform algorithms have been used for texture classification, such as Fourier Transform, wavelet transform and Gabor filters.

Gabor filter is one of the most established texture descriptors introduced by Gabor in 1946 [6]. It is applied to extract features by analyzing the frequency domain of the image .Gabor filter is basically a Gaussian function modulated by complex sinusoidal of frequency and orientation. It has the ability to perform both in spatial and frequency domain. Therefore Several studies in [7,8] has been shown Gabor filter for texture feature extraction to outperform other previous transform based methods. One of the most popular multiresolution analysis tools is the wavelet transform. In wavelet analysis an image is usually decomposed at different scales and orientations with localization in both time and frequency [10].

Discrete wavelet transform (DWT) can be implemented as a set of high-pass and low-pass filter banks. In standard wavelet decomposition, the output from the low-pass filter can be then decomposed further, with the process continuing recursively.

In the last decade, wavelet transform has been recognized as a powerful tool in a wide range of applications, including image processing, compression, denoising, and classification. Wavelet transform has been employed to extract features from magnetic resonance imaging in [13,14]. To achieve higher recognition rate using wavelet features, various methods have been developed .

In [15] the authors utilize stationary wavelet transform (SWT) to extract features instead of DWT. traditional discrete wavelet transform (DWT) suffers from translation variant property, which may extract significantly different features from two images of the same subject with only slight movement.

Even though wavelet transform is an effective tool for feature extraction, but it requires large storage and is computationally more expensive . Hence an alternative method for dimension reduction scheme is used. In order to reduce the feature vector dimension and increase the discriminative power, the principal component analysis (PCA) [16] has been used.

In [17] the proposed system makes use of PCA for reduction of dimensionality of the feature space and therefore reduces the computational cost of analyzing new data. The purpose here is to obtain an optimal subset of features which produce the best possible results.

In [18] another feature selection method, linear Discriminant analysis (LDA) that is closely related to PCA. Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performance has been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. therefore texture features were extracted from MRI Brain images using wavelet and then vital features are selected using LDA for better classification rate.

In [19,20] the authors utilize genetic algorithm (GA) as a feature selection method for wavelet extracted features, GA is an artificial intelligence procedure based on the theory of natural selection and evolution . It is an efficiently global searching algorithm based on the principle of survival of the fittest and used for optimization and searching problems.

Many recent developments in multiresolution analysis (MRA) have taken place, while wavelets are suitable for dealing with objects with point singularities. Wavelets capture limited directional information due to its poor orientation selectivity. By decomposing the image into a series of high-pass and low-pass filter bands.

The wavelet transform extracts directional details that capture horizontal, vertical, and diagonal activity. However, these three linear directions are limiting and might not capture enough directional information in noisy images, such as medical images, Ridgelet improves MRA segmentation; however, they capture structural information of an image based on multiple radial directions in the frequency domain. Line singularities in ridgelet transform provides better edge detection than its wavelet counterpart. One limitation to use ridgelet in image segmentation is that ridgelet is most effective in detecting linear radial structures, which are not dominant in medical images[21].

Curvelet transform is a multi-scale and multi-directional transform developed by Candes and Donoho [22,23] to overcome the limitations of the conventional wavelet transform. Curvelet transform provides almost optimal representation of objects with curve singularities. That means, curvelet transform needs relatively a small number of coefficients to represent a line or a curve in a given medical images, It is very important to detect and to enhance the boundaries between different structures in medical image processing.

In [24], the curvelet transform is applied on a set of texture images. One group feature vector can be constructed by the mean and variance of the Curvelet Statistical Features (CSFs), which are derived from the sub-bands of the curvelet decomposition and are used for classification.

In [25], the authors offer a curvelet transform approach for the fusion of Magnetic Resonance and Computed Tomography. Curvelet transform provides good results in their fusion experimentation.

Eltoukhy et al. [26] present an approach for diagnosis of breast cancer via curvelet transform. Their feature vector includes biggest coefficients from curvelet transform. They select a ratio (10%: 90%) of features for every image and they classify the selected part using the Euclidean distance classifier. They use 142 mammogram images from MIAS data set.

Eltoukhy et al. [27] introduce a study of mammogram classification based on curvelet transform. They decomposed each image into different scales and angles. They select different ratio from biggest coefficients to use as feature vector and they use different decomposition level to detect the scale that provides maximum classification accuracy. Maximum success rate is obtained 98.59% for normal and abnormal classification at scale 2 with 40% of the biggest coefficients, and at scale 3 with 40% and 70% of the biggest coefficients. Their comparative study with wavelet show that features from curvelet yield a greater classification success rate than wavelet based features.

Dehghani et al. [28] propose a system based on texture feature extraction and SVM. Feature extraction is done by using statistical method and signal processing method from ROIs. The system is used a new classifier based on SVM called Weighted SVM. They compare the features from Co-occurrence matrix, Wavelet transform, Contourlet transform Co-occurrence matrix by using SVM and Weighted SVM.

2.3 Classification

Machine learning algorithms are described as either 'supervised' or 'unsupervised'. The distinction is drawn from how the learner classifies data. In supervised algorithms, the classes are predetermined. These classes can be conceived of as a finite set, previously arrived at by a human. In practice, a certain segment of data will be labeled with these classifications. The machine learner's task is to search for patterns and construct mathematical models. These models then are evaluated on the basis of their predictive capacity in relation to measures of variance in the data itself. Unsupervised learners are not provided with classifications. In fact, the basic task of unsupervised learning is to develop classification labels automatically. Unsupervised algorithms seek out similarity between pieces of data in order to determine whether

they can be characterized as forming a group. These groups are termed clusters, and there are a whole family of clustering machine learning techniques.

In unsupervised classification, often known as 'cluster analysis' the machine is not told how the texts are grouped. Its task is to arrive at some grouping of the data [29].

2.3.1 Supervised Classifications

Supervised classification is one of the most frequently decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object.

The application of machine learning techniques to medical image analysis and pattern recognition problems has gained a widespread acceptance. Among the techniques developed for classification, Bayes classifier, Decision Trees, K-Nearest Neighbors (KNN) classifier, Artificial Neural Networks (ANN) Support Vector Machines (SVMs) and Expectation Maximization(EM) as a statistical classification scheme[30].

Artificial neural network deployment (ANN) excel in the supervised techniques for brain tissues classification into cancerous or non-cancerous automatically. In [31] A new approach for automated diagnosis and classification of Magnetic Resonance (MR) human brain images is proposed. The proposed method uses discrete wavelet for feature extraction, employs PCA for feature selection and exploits the capability of Back propagation neural network (BPN) and radial basis function neural network (RBFN) to classify brain MRI images automatically.

In [32] Probabilistic Neural Network (PNN) was applied for classification. It is adopted because training speed is many times faster than a Back propagation neural network (BPN), The performance of the PNN classifier was evaluated in terms of training performance and classification accuracies. Probabilistic Neural Network gives fast and accurate classification and is a promising tool for classification of the tumors. The observed results are significantly better than the results reported in researches employing Wavelet Transform and other classification methods.

In [34] The authors use a popular classification technique, Decision Trees due to their ability to model non-linear dependencies in the features, and their intuitive graphical representation of the learned model (as opposed to, for example, ANNs). Decision Trees perform classification by making a set of decisions based on the features, beginning from a root 'node' and following decision made to other nodes in the tree where new decisions are made, leading finally to a 'leaf' where a classification is made.

In [35] Support vector machines are a pattern recognition technique grown up from statistical learning theory. The basic idea of applying SVMs for solving classification problems is to transform the input space to higher dimension feature space through a non-linear mapping function then it constructs the separating hyperplane with maximum distance from the closest points of the training set.

In the case of linear separable data, the SVM tries to find among all hyper planes that minimize the training error, the one that separates the training data with maximum distance from their closest points.

In [36], Sachdeva et al. developed a CADx system for assisting radiologists in multiclass classification of brain tumors. A new hybrid machine learning system based on the Genetic Algorithm (GA) and Support Vector Machine (SVM) for brain tumor classification is proposed. Texture and intensity features of tumors are taken as input. Genetic algorithm has been used to select the set of most informative input features.

In [37], a new improved method is implemented by combining LDA & PCA for feature reduction and SVM is used for classification of MRI images. Compared to the previous work suggested in the literature discussed above high accuracy is achieved for feature selection and extraction.

Artificial Neural Networks and Fuzzy systems are the widely preferred a techniques for biological computational applications. While ANN is less accurate than fuzzy logic systems, fuzzy theory needs expertise knowledge to guarantee high accuracy. Since both the methodologies possess certain advantages and disadvantages, it is

important to integrate these two techniques in one hybrid tumor classification system.

In [44] Joshi et al. proposed a brain cancer detection and classification System using Neuro Fuzzy logic classifier. Texture features are used in the training of the artificial neural network.   Co-occurrence matrices at different directions are calculated and Grey Level Co-occurrence Matrix (GLCM) features are extracted from the matrices.

2.3.2 Unsupervised clustering

Image segmentation refers to the process of partitioning a digital image into multiple regions (set of pixels) [39]. The goal of segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in an image.

Image segmentation requires extracting specific features from an image by distinguishing objects from the background. The process involves classifying each pixel of an image into a set of distinct classes, where the number of classes is much smaller. Medical image segmentation aims to separate known anatomical structures from the background such cancer diagnosis, quantification of tissue volumes, radiotherapy treatment planning, and study of anatomical structures.

Segmentation can be manually performed by a human expert who simply examines an image, determines borders between regions, and classifies each region. This is perhaps the most reliable and accurate method of image segmentation, because the human visual system is immensely complex and well suited to the task. But the limitation starts in volumetric images due to the quantity of clinical data.

K-means algorithm which Place K points into the space represented by the objects that are being clustered. These points represent initial group of centroids. Assign each object to the group that has the closest centroids, When all objects have been assigned, recalculate the positions of the K centroids. Repeat until the centroids no

longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

In [38], A proposed system explains how to extract the brain tumor in patient's body using s K-means as clustering algorithm. In the second step they perform classification of MRI brain image using decision tree and SVM classification algorithms and predict which is better classification technique.

Fuzzy C-means (FCM) is a method of clustering which allows one pixel to belong to two or more clusters [39]. The FCM algorithm attempts to partition a finite collection of pixels into a collection of "c" fuzzy clusters with respect to some given criterion. Depending on the data and the application, different types of similarity measures may be used to identify classes. Some examples of values that can be used as similarity measures include distance, connectivity, and intensity. In [40] A modified FCM algorithm is used for clustering abnormal MR brain images from four tumor classes namely metastase, meningioma, glioma and astrocytoma. The segmented outputs are analyzed based on the segmentation efficiency and convergence rate.

A number of approaches have been used to segment and predict the grade and volume of the brain tumor. In [42] EI papageevgious et.al (applied soft computing 2008) in their work proposed a fuzzy cognitive map (FCM) to find the grade value of brain tumor.

In [43] A new approach for automated detection of brain tumor based on k-means and possibilistic c-means clustering with color segmentation, which separates brain tumor from healthy tissues in magnetic resonance images. The magnetic resonance feature images used for the tumor detection consist of T1-weighted and T2-weighted images for each axial slice through the head. The proposed method consists of three stages namely pre-processing, segmentation and feature extraction.

In [44] Computer-Aided Diagnosis system with watershed segmentation for Automatic detection of brain tumor through MRI was developed. The concept of watersheds is based on visualizing an image in three dimensions, two spatial coordinates versus grey levels. In such a topographic interpretation, As watershed segmentation technique segregates any image as different intensity portions and also the tumerous cells have high proteinaceous fluid which has very high density and

hence very high intensity, therefore watershed segmentation is the best tool to classify tumors and high intensity tissues of brain.

In [45] Implementation of a neuro-fuzzy segmentation process of the MRI data is presented in this study to detect various tissues like white matter, gray matter, csf and tumor. The advantage of hierarchical self organizing map and fuzzy c means algorithms are used to classify the image The performance of the ANFIS classifier was evaluated in terms of training performance and classification accuracy. Here the result confirmed that the proposed ANFIS classifier with accuracy greater than 90 percent has potential in detecting the tumors. In [50] the authors propose an automatic and accurate technique for classifying normal and abnormal magnetic resonance (MR) images of human brain. Ripplet transform Type-I (RT), an efficient multiscale geometric analysis (MGA) tool for digital images, is used to represent the salient features of the brain MR images. The dimensionality of the image representative feature vector is reduced by principal component analysis (PCA). A computationally less expensive support vector machine (SVM), called least square-SVM (LS-SVM) is used to classify the brain MR images.

Chapter 3
Background

Building a completely automatic and efficient medical image diagnosis system is a grand challenge for biometrics, computer vision and pattern recognition researchers. Generally, such a recognition system is able to perform three subtasks: feature extraction, feature reduction, and classification. In this thesis we will put our focus on feature extraction, the significant step prior to classification. The key issue here is to construct a representative feature set that can enhance system-performance both in terms of accuracy and speed. Most of the feature extraction algorithms aim at effectively representing of the extracted information. Some of these algorithms are based on either the Fourier transform or the wavelet transform. Unfortunately, Neither the Fourier transform nor the wavelet transform can represent the directional features accurately. Curvelet Transform (CT) is a geometric transform developed by Emmanuel Cand`es et al. [22] to overcome these inherent limitations.

3.1 Curvelet Transform

Curvelet transform is a multi-scale and multi-directional transform with needle shaped basis functions which has high directional sensitivity and anisotropy. Curvelets obey parabolic scaling law: width$\cong$(length)^2. Therefore, curvelet transform allows almost optimal sparse representation of curve singularities The curvelet transform at different scales and directions span the entire frequency space, which is not the case with other directional transforms such as Gabor and wavelets. Curvelets are designed to represent edges and other singularities along curves much more efficiently than the traditional Wavelet transform which is good at representing point singularities. Figure 3.1 shows edge representation by both Wavelets and Curvelet Transforms. It can be noticed, it would take many Wavelet coefficients to accurately represent such a curve while Curvelet needs small number of coefficients; wavelet needs three, six, and twelve coefficients, while Curvelet needs one, two, and four coefficients, in the largest, middle, and smallest scale respectively.
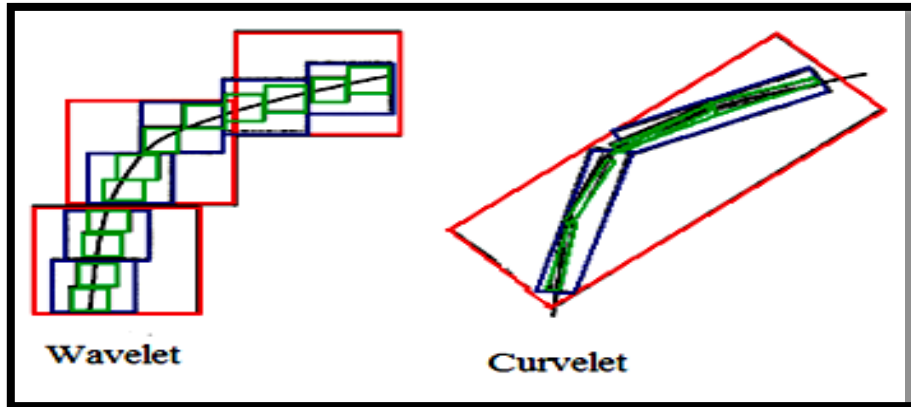
Figure 3.1: Edge Representations using wavelet and ridgelet.

To explain how the Curvelet basis elements align with edges in an image and how this alignment affects the coefficients of the corresponding transform. Figure 3.2 pictorially depicts the alignment. The first box shows the original image. In the second box, the band-passed image (image at a certain resolution) is shown. Finally in the third box the alignment of the Curvelet basis elements with a small section of the edge is shown and will now be briefly explained.

Figure 3.2: Curvelet Alignments.

The frequency tiling by the discrete curvelet transform is shown in Fig. 3.3. We can observe that the entire frequency plane is covered by the curvelets at different scales and orientations.

Figure 3.3: Frequency tiling of whole image by curvelet transform in discrete domain and support of a curvelet at scale *j* in frequency domain showing its length and width.

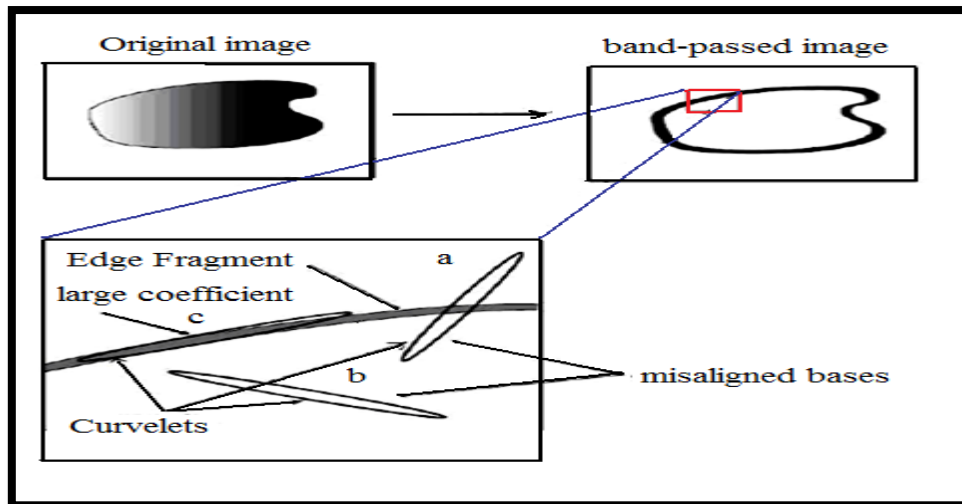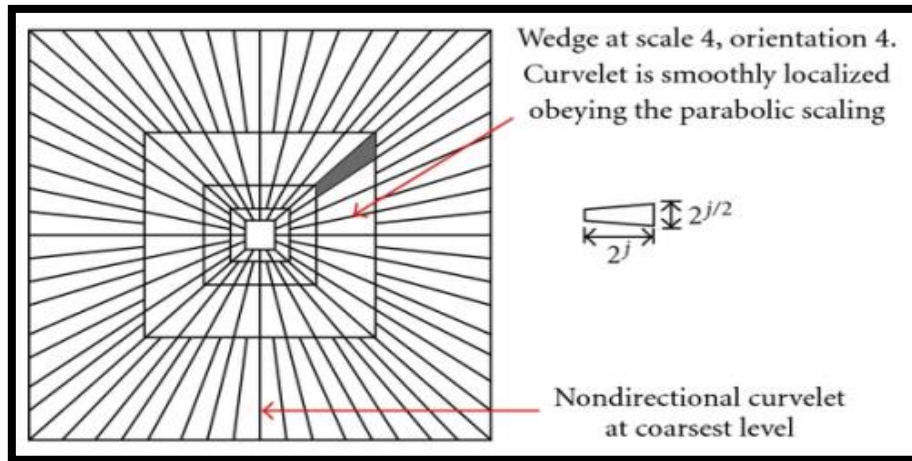Curvelet transform has undergone a major revision since its invention [22]. The first generation curvelet transform is based on the concepts of ridgelet transform [22]. The curve singularities have been handled by smooth partitioning of the bandpass images. In each smooth partitioned block the curve singularities can be approximated to a line singularity. A ridgelet transform is applied on these small blocks, where ridgelets can deal the line singularities effectively. To avoid blocking artifacts, the smooth partitioning is done on overlapping blocks which results in redundancy, and the whole process involves subband decomposition using wavelet transform, smooth partitioning and ridgelet analysis on each block; this process consumes more time. The implementation of second generation curvelet transform is based on the Fourier transform and is faster, less complex, and less redundant [22]. But ridgelet based curvelet transform is not efficient as it uses complex ridgelet transform [23]. In 2005, Candès et al. proposed two new forms of curvelet transform based on different operations of Fourier samples [26], namely, unequally-spaced fast Fourier transform (USFFT) and wrapping based fast curvelet transform. Wrapping based curvelet transform is faster in computation time, simple, easy to understand, and more robust than ridgelet and USFFT based curvelet transform [27]. To our knowledge, wrapping based curvelet transform has not been used in computer aided diagnosis systems and there is no work on a systematic evaluation of curvelet in CADx systems.

Fig. 3.4 shows some curvelets at different scales and orientations in both spatial and frequency domain. As the scale is increased, the curvelets are more sensitive to the

orientation. This can be observed in this figure. The curvelet in spatial domain at higher scale is smaller and finer to the lower one and it can be observed that curvelets show oscillating behavior perpendicular to the orientation in the frequency domain.

Basically, curvelet transform is a multiscale transform with a pyramid structure consisting of many orientations at each scale. This pyramid structure consists of several subbands at different scales in the frequency domain. Subbands at high and low frequency levels have different orientations and positions. At high scales, the curvelet waveform becomes so fine that it looks like a needle shaped element (left images of Figure 3.4 (e), (f)). Whereas, the curvelet is non directional at the coarsest scale (left image of Fig. 3.4(a)) [26]. With increase in the resolution level the curvelet becomes finer and smaller in the spatial domain and shows more sensitivity to curved edges which enables it to effectively capture the curves in an image (Figure. 3.4). As a consequence, curved singularities can be well-approximated with few coefficients. High frequency components of an image play a vital role in finding distinction between images. Curvelets at fine scales effectively represent edges by using texture features computed from the curvelet coefficients. Curvelets at different scales and their frequency responses are shown in Fig. 3.4. These are generated using Curvelab-2.1.2 of [46].

Discrete curvelet transform of a two dimensional image $f[m.n](0 \leq m \leq M-1, 0 \leq n \leq N-1)$ at a given scale, orientation, and position is given by equation 3.1 bellow:

$$C^D(j,l,k) = \sum_{\substack{0 \leq m < M, \\ 0 \leq n < N}} f[m.n] \overline{\varphi^D_{J,l,k}[m,n]} \tag{3.1}$$

where $C^D(j,l,k)$ is the curvelet coefficient at scale j, orientation l, and position $k[k_1,k_2]$. In the above equation $\varphi^D_{j,l,k}[m,n]$ is the digital curvelet waveform. In the frequency domain the mother curvelet is represented as a product of two windows, radial and angular windows as shown in equation 3.2 given bellow

$$\widetilde{U_{j,l}}(w) = \widetilde{W_j}(w)V_j(w) \tag{3.2}$$

In the above equation $V_j(w)$ is a smooth, real valued, and non negative angular window obeying the admissibility conditions and $\widetilde{W_j}(w)$ is the Cartesian equivalent of radial window, supported on concentric squares and is given equation 3.3 bellow:
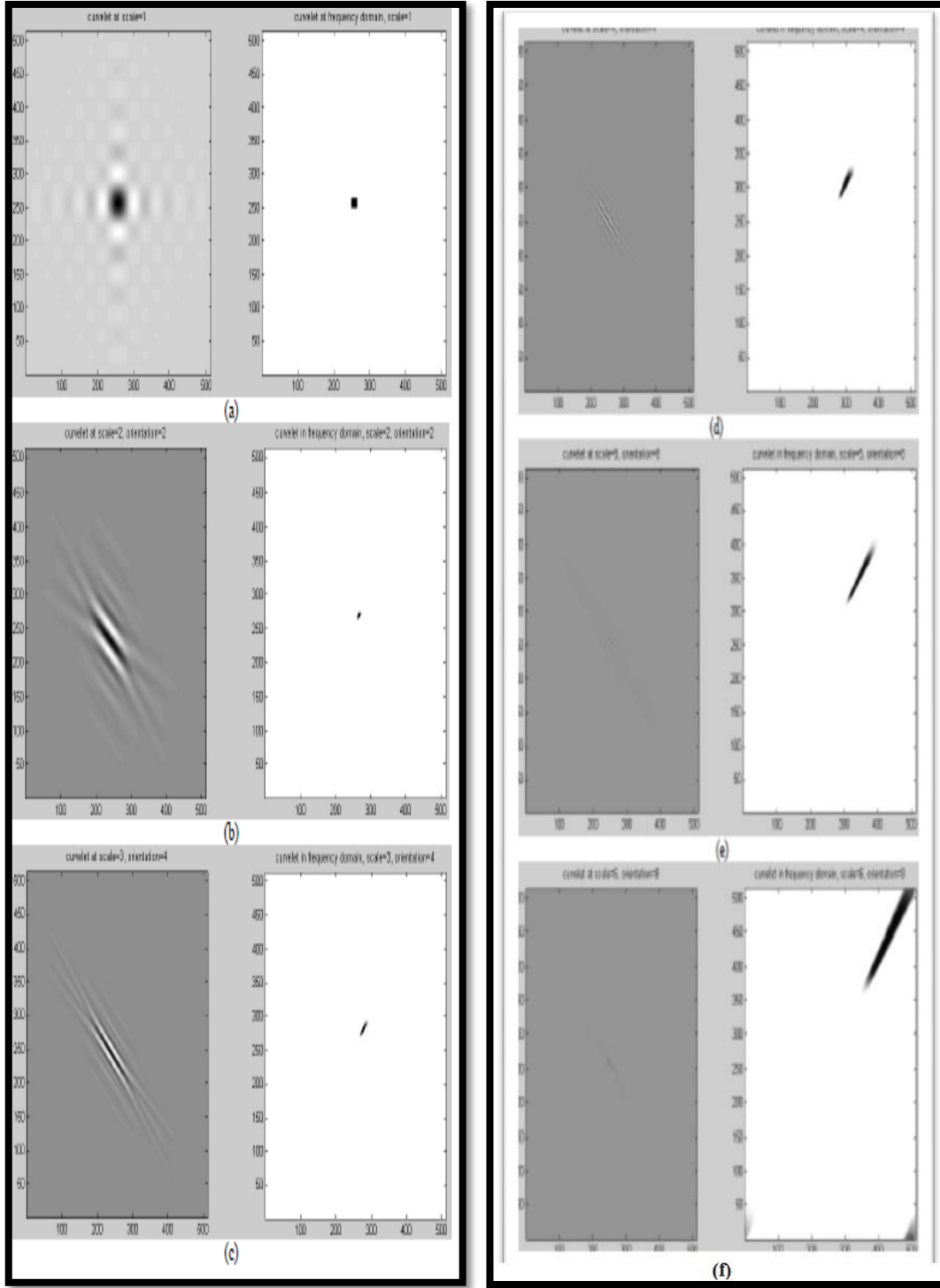
Figure 3.4: Curvelets (absolute value) at different scales, directions are shown in the spatial domain (left) and in the frequency domain (right).

$$\widetilde{W_J}(w) = \sqrt{\emptyset_{j+1}^2 - (w) - \emptyset_j^2(w)}, \, j \geq 0 \qquad (3.3)$$

Where $\emptyset$ is defined as the product of low pass one dimensional windows given equation 3.4 below:

$$\emptyset_j(w_1, w_2) = \emptyset(2^{-1} w1)\emptyset(2^{-j} w_2) \tag{3.4}$$

Function $\emptyset$ is equal to 1 in [-1/2,1/2] and vanishes outside [-2,2].Using the window $\widetilde{W_j}(w)$ we can separate the scales of the frequency plane. The angular localization is obtained using $V_j(w)$The product of $V_j(w)$ and $\widetilde{W_j}(w)$ isolates the frequencies near the wedge $2^j \leq w_1 \leq 2^{j+1}, -2^{-j/2} \leq w_2 \leq 2^{j/2}$ [13]. In the frequency domain in curvelet coefficient can be obtained as :

$$Curvelet\ Coefficient = IFFT\ [FFT(Curvelet)X\ FFT\ (Image)]$$

where IFFT is the inverse fast Fourier transform and FFT is the fast Fourier transform. The digital curvelet waveform in frequency domain is obtained by the product of two windows, called radial window and angular window. The support of the wedge like digital curvelet waveform is not rectangle and hence IFFT cannot be applied to the product to find the curvelet coefficient. A wrapping technique has been developed by Cand`es et al. [22] to solve this problem. The idea behind the wrapping technique is to select a parallelogram that can support the wedge shaped digital curvelet waveform and wrap the parallelogram around the origin to obtain a rectangular support to apply IFFT. The wrapping technique is explained in Figure. 3.5.
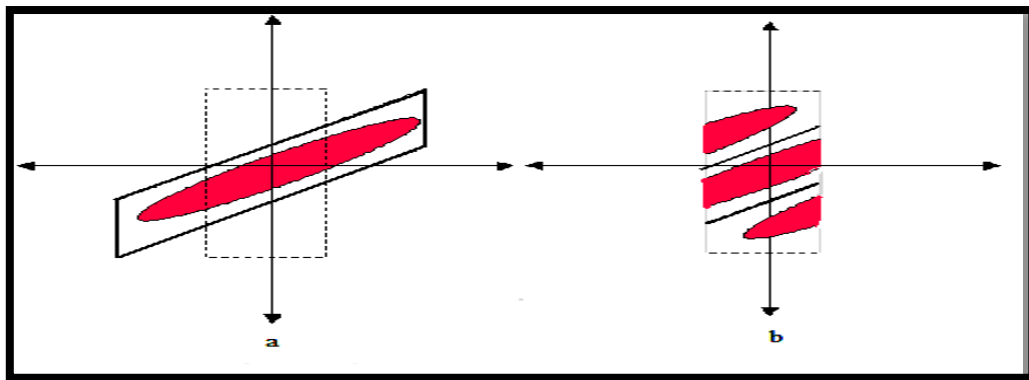

Figure 3.5: Wrapping method. The support in a parallelogram is finally into a rectangle (Quoted from [22]).

The periodic tiling of the parallelogram results in a rectangular support at the center on which IFFT is obtained. The wrapping based curvelet transform is calculated as follows [13]:
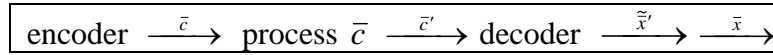
a. Apply the 2D FFT and obtain Fourier samples $\hat{f}[m, n]$ of the image.

b. Find the product of Fourier domain digital curvelet waveform and the Fourier samples of the image.

c. Wrap the product around the origin to obtain rectangular support.

d. Apply the inverse 2D FFT on the wrapped product.

The contrast between wavelet and ridgelet on capturing edge information is shown in Fig. 3.4. It can be observed that the curvelets, at all scales, capture the edge information more accurately and tightly than wavelets.

3.2 Feature Dimensionality Reduction

At the core of computer aided diagnosing system is the extraction of proper features. Direct use of pixel values as features is not possible due to huge dimensionality of the brain images. A common problem in data processing is that large amounts of data are expensive to be processed. To reduce the amount of data, would mean a reduction in expenses. But simply throwing away part of the data would result in a loss of information, which could be important. In any random data, like for example data from images, there is a variation in how important each part of data is to the information which is stored in the data. By leaving out the part of data which is the least valuable to the information, we reach a reduction of the amount of data. Traditionally, Principal Component Analysis (PCA) is employed to obtain a lower dimensional representation of the data. PCA is used to compress data in such a way that the least information is lost. It does so by truncating data and thereby leaving out the data which is of the least importance to the information stored in the data. This PCA process is called dimensionality reduction, because a vector $\bar{x}$ which contains the original data and is N-dimensional is reduced to a compressed vector $\bar{c}$ which is M-dimensional, where M<N. The question that is answered by PCA is: how can we map vector $\bar{x}$ into a vector $\bar{c}$ with a smaller dimension, but where the information contained in $\bar{x}$ is equal or less equal to the information stored in $\bar{c}$. So, which linear

operation should be performed on vector $\bar{x}$ to transform it to vector $\bar{c}$, where information $\bar{x} \sim$ information $\bar{c}$, and dim $\bar{c} <$ dim $\bar{x}$?. We can visualize the above in the following diagram.

$$\boxed{\text{encoder} \xrightarrow{\bar{c}} \text{process } \bar{c} \xrightarrow{\bar{c}'} \text{decoder} \xrightarrow{\tilde{\tilde{x}}'} \xrightarrow{\bar{x}}}$$

A vector $\bar{x}$ is coded into a vector $\bar{c}$ with a reduced dimension. Vector $\bar{c}$ is then processed, which results in vector $\bar{c}'$, which can be decoded back to a vector $\tilde{\tilde{x}}'$. This last vector is an approximation of the result which would have been attained by storing, transmitting or processing vector $\bar{x}$.

- **Mathematics of Principal Components**

We start with p-dimensional feature vectors, and want to summarize them by projecting down into a L-dimensional subspace. Our summary will be the projection of the original vectors on to q directions, the principal components, which span the sub-space. There are several equivalent ways of deriving the principal components mathematically. The simplest one is finding the projections which maximize the variance. The first principal component is the direction in feature space along which projections have the largest variance. The second principal component is the direction which maximizes variance among all directions orthogonal to the first. The Jth component is the variance-maximizing direction orthogonal to the previous J-1 components. There are p principal components in all. The following is a detailed description of PCA using the covariance method :

- **Step1: Organize the data set**

Suppose you have data comprising a set of observations of p variables, and you want to reduce the data so that each observation can be described with only L variables, L< p. Suppose further, that the data are arranged as a set of n data vectors (X1….Xn) with each Xi representing a single grouped observation of the p variables. Write (X1 …. Xn) as row vectors, each of which has p columns. Then place the row vectors into a single matrix X of dimensions n x p.

- **Step2: Calculate the empirical mean**

Find the empirical mean along each dimension j = 1, ..., p. Then place the calculated mean values into an empirical mean vector u of dimensions p × 1.

$$u[j] = \frac{1}{N}\sum_{i=1}^{N} X[i,j].$$

(3.5)

- **Step3: Calculate the deviations from the mean**

Mean subtraction is an integral part of the solution towards finding a principal component basis that minimizes the mean square error of approximating the data. Hence we proceed by centering the data as follows:

Subtract the empirical mean vector **u** from each row of the data matrix **X**. Store mean-subtracted data in the $n \times p$ matrix **B**.

$$B = X - hu^T$$

(3.6)

where **h** is an $n \times 1$ column vector of all 1s:

$$h[i] = 1 \qquad\qquad for\ i = 1, \ldots, n$$

(3.7)

- **Step4: Find the covariance matrix**

    Find   the $p \times p$ empirical covariance   matrix **C** from   the outer   product of matrix B with itself:

(3.8)

$$C = \frac{1}{N-1}\ B^* \cdot B$$

Where * is the conjugate transpose operator. Note that if B consists entirely of real numbers, which is the case in many applications, the "conjugate transpose" is the same as the regular transpose.

- **Step5: Find the eigenvectors and eigenvalues of the covariance matrix**

    Compute   the   matrix V of eigenvectors which diagonalizes the   covariance matrix C:

$$V^{-1}\ CV = D$$

(3.9)

where $\mathbf{D}$ is the diagonal matrix of eigenvalues of $\mathbf{C}$. This step will typically involve the use of a computer-based algorithm for computing eigenvectors and eigenvalues. Matrix $\mathbf{D}$ will take the form of an $M \times M$ diagonal matrix, where

$$D[k, l = \lambda_k \quad for \ k = l = j \tag{3.10}$$

is the jth eigenvalue of the covariance matrix $\mathbf{C}$, and

$$D[k, l] = 0 \quad for \ k \neq l \tag{3.11}$$

Matrix $\mathbf{V}$, also of dimension $p \times p$, contains p column vectors, each of length p, which represent the p eigenvectors of the covariance matrix $\mathbf{C}$. The eigenvalues and eigenvectors are ordered and paired. The jth eigenvalue corresponds to the jth eigenvector.

- **Step 6: Rearrange the eigenvectors and eigenvalues**

  Sort the columns of the eigenvector matrix $\mathbf{V}$ and eigenvalue matrix $\mathbf{D}$ in order of decreasing eigenvalue.

$$g[j] = \sum_{k=1}^{j} d[k, k] \ for \ j = 1, \ldots, p \tag{3.12}$$

- **Step 7: Select a subset of the eigenvectors as basis vectors**

  Save the first L columns of $\mathbf{V}$ as the $p \times L$ matrix $\mathbf{W}$:

$$W[k, l] = V[k, l] \quad for \ k = 1, \ldots, p \quad l = 1, \ldots, K \tag{3.13}$$

Where $\quad 1 \leq L \leq p$.

Use the vector $\mathbf{g}$ as a guide in choosing an appropriate value for L. The goal is to choose a value of L as small as possible while achieving a reasonably high value of g on a percentage basis. For example, you may want to choose L so that the cumulative energy g is above a certain threshold, like 90 percent. In this case, choose the smallest value of L such that

$$\frac{g[L]}{g[p]} \geq 0.9 \tag{3.14}$$

3.3 Artificial Neural Networks

Classification is a data mining (machine learning) technique used to predict group membership for data instances. To simplify the problems of prediction or classification, neural networks are being introduced. Neural networks are simplified models of the biological neuron system. It is an extremely parallel distributed processing system made up of highly interconnected neural computing elements that have the ability to learn and thereby acquire knowledge and make it available for use. Various learning mechanisms exist to enable the NN acquire knowledge. NN architectures have been classified into various types based on their learning mechanisms and other features.

This learning process is referred to as training and the ability to solve a problem using the knowledge acquired as inference. NNs are simplified imitations of the central nervous system[10], and therefore, have been inspired by the kind of computing performed by the human brain. The structural constituents of a human brain termed neurons which perform computations such as cognition, logical inference, pattern recognition. Hence the technology, which has been built on a simplified imitation of computing by neurons of a brain, has been termed Artificial Neural Networks (ANN) or simply neural networks. Computers, Parallel Distributed Processors etc. Also, neurons are can also Processing Elements (PEs), and nodes.

Neural network Advantages:

- A neural network can perform tasks that a linear program cannot.
- When an element of the neural network fails, it can continue without any problem by their parallel nature.
- A neural network learns and does not need to be reprogrammed.
- It can be implemented in any application.

Disadvantages:

- The neural network needs training to operate.
- The architecture of a neural network is different from the architecture of microprocessors therefore needs to be emulated.
- Requires high processing time for large neural networks.

3.3.1 Properties of Neural Networks

- The NNs display mapping capabilities, that can map input patterns to their associated output patterns.[10]

- The NNs learn by examples. Thus, NN architectures can be trained with known examples of a problem before they are tested for their inference capability on unknown instances of the problem. They can, therefore, identify new objects previously untrained.

- The NNs possess the ability to generalize. Thus, they can predict new outcomes from the past trends.

- The NNs are robust systems and are fault tolerant. They can, therefore, recall full patterns from incomplete, partial or noisy patterns.

- The NNs can process information in parallel, at high speed, and in a distributed manner.

3.3.2 Neural Networks Characteristics

The word network in Neural Network refers to the interconnection between neurons present in various layers of a system. Every system is basically a 3 layered system, which are Input layer, Hidden Layer and Output Layer. The input layer has input neurons which transfer data via synapses to the hidden layer, and similarly the hidden layer transfers this data to the output layer via more synapses. The synapses stores values called weights which helps them to manipulate the input and output to various layers. An ANN can be defined based on the following three characteristics:

1. The Architecture: The number of layers and the no. of nodes in each of the layers
2. The learning mechanism has been applied for updating the weights of the connections.
3. The activation functions used in various layers.

3.3.3 Functions of Neural networks

Artificial neural networks (ANN) serve two important functions: as pattern classifiers and as nonlinear adaptive filters. An Artificial Neural Network is an adaptive, most often nonlinear system that learns to perform a function (an input/output map) from data. Adaptive neural network means that the system parameters are changed during operation, normally called the training phase. After

the training phase the Artificial Neural Network parameters are fixed and the system is deployed to solve the problem at hand (the testing phase).

The Artificial Neural Network is built with a systematic step-by-step procedure to optimize a performance criterion or to follow some implicit internal constraint, which is commonly referred to as the learning rule. The input/output training data are fundamental in neural network technology, because they convey the necessary information to discover the optimal operating point. The nonlinear nature of the neural network (PEs) provides the system with lots of flexibility to achieve practically any desired input/output map, i.e., some Artificial Neural Networks are universal mappers .
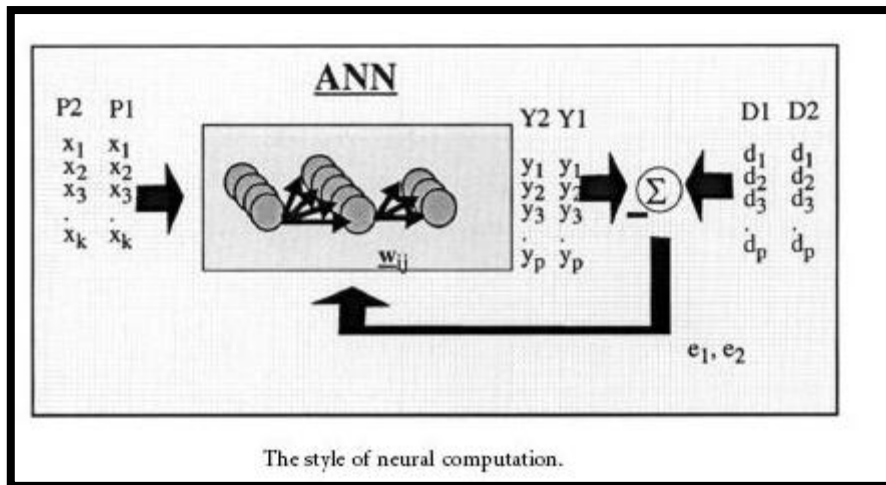

Figure 3.6 Approach of Neural Computing.

In figure 3.6 an input is presented to the neural network and a corresponding desired or target response set at the output. An error is composed from the difference between the desired response and the system output. This error information is fed back to the system and adjusts the system parameters in a systematic fashion (the learning rule). The process is repeated until the performance is acceptable. It is clear from this description that the performance hinges heavily on the data. If one does not have data that cover a significant portion of the operating conditions or if they are noisy, then neural network technology is probably not the right solution. On the other hand, if there is plenty of data and the problem is poorly understood to derive an approximate model, then neural network technology is a good choice.

3.3.4 ANN Technology

This operating procedure should be contrasted with the traditional engineering design, made of exhaustive subsystem specifications and intercommunication protocols. In artificial neural networks, the designer chooses the network topology, the performance function, the learning rule, and the criterion to stop the training phase, but the system automatically adjusts the parameters. So, it is difficult to bring a priori information into the design, and when the system does not work properly it is also hard to incrementally refine the solution. But ANN-based solutions are extremely efficient in terms of development time and resources, and in many difficult problems artificial neural networks provide performance that is difficult to match with other technologies.

3.3.5 Neuron Model

Simple Neuron

A neuron with a single scalar input and no bias appears on the left below.
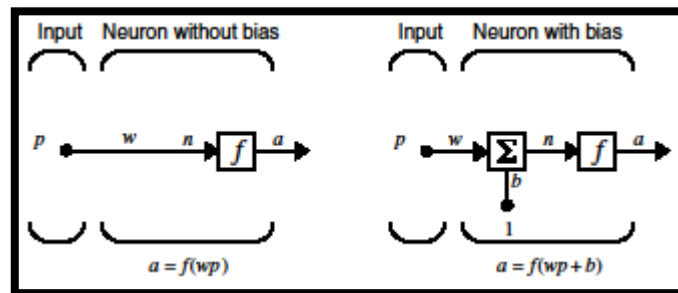
Figure 3.7 Neuron Model.

The scalar input p is transmitted through a connection that multiplies its strength by the scalar weight w to form the product wp, again a scalar. which produces the scalar output a. The neuron on the right has a scalar bias, b. You can view the bias as simply being added to the product wp as shown by shifting the function f to the left by an amount b. The bias is much like a weight, except that it has a constant input of 1. The transfer function net input n, again a scalar, is the sum of the weighted input wp and the bias b. The central idea of neural networks is that such parameters can be adjusted so that the network exhibits some desired or behavior.

Neuron with Vector Input

A neuron with a single R-element input vector is shown below. Here the individual element inputs P_1, P_2, ----P_R and $w_{1,1}$, $w_{1,2}$, ---$w_{1,R}$ are multiplied. Then the weighted values are fed to the summing junction. Their sum is simply **Wp**, the dot product of the (single row) matrix **W** and the vector **p**.
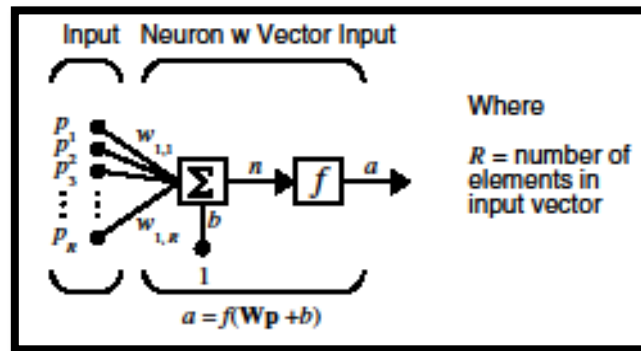
Figure 3.8 Neuron With Vector Input

The neuron has a bias b, which is summed with the weighted inputs to form the net input n. This sum, n, is the argument of the transfer function f.

3.3.6 Network Architectures

Two or more of the neurons shown earlier can be combined in a layer, and a particular network could contain one or more such layers. First consider a single layer of neurons.

A Layer of Neurons

A one-layer network with R input elements and S neurons follow. In this network, each element of the input vector **p** is connected to each neuron input through the weight matrix **W**. The i_{th} neuron has a summer that gathers its weighted inputs and bias to form its own scalar output $n(i)$. The various $n(i)$ taken together form an S-element net input vector **n**. Finally, the neuron layer outputs form a column vector **a**. The expression for **a** is shown at the bottom of the figure.
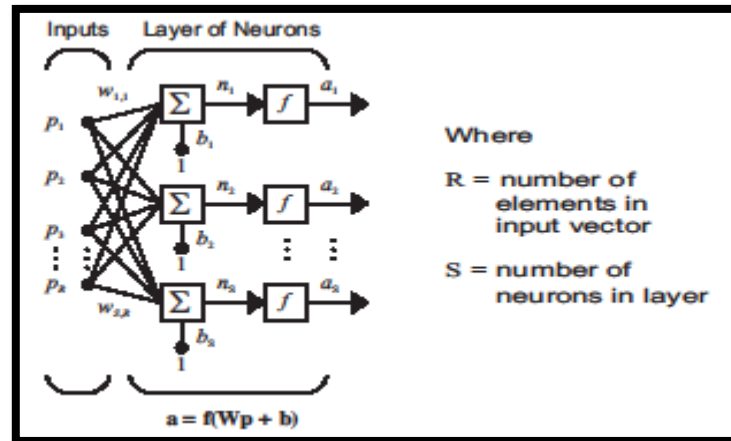
Figure 3.9 Layer of Neurons.

A layer is not constrained to have the number of its inputs equal to the number of its neurons. You can create a single (composite) layer of neurons having different transfer functions simply by putting two of the networks shown earlier in parallel. Both networks would have the same inputs, and each network would create some of the outputs.
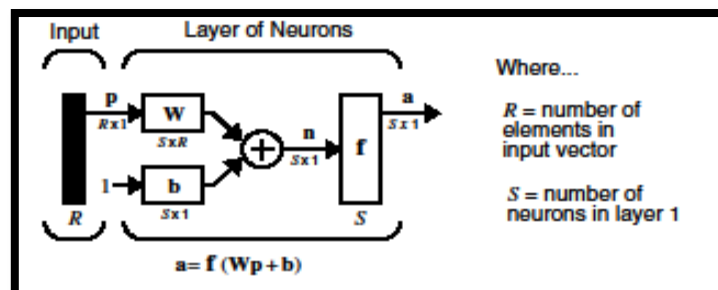
Figure 3.10 : Composite Layer of Neurons.

Here **p** is an R length input vector, **W** is an SxR matrix, and **a** and **b** are S length vectors. As defined previously, the neuron layer includes the weight matrix, the multiplication operations, the bias vector **b**, the summer, and the transfer function boxes.

3.3.7 The Mathematical Model

When creating a functional model of the biological neuron, there are three basic components of importance. First, the synapses of the neuron are modeled as weights.

The strength of the connection between an input and a neuron is noted by the value of the weight. Negative weight values reflect inhibitory connections, while positive values designate excitatory connections [Haykin]. The next two components model the actual activity within the neuron cell. An adder sums up all the inputs modified by their respective weights. This activity is referred to as linear combination. Finally, an activation function controls the amplitude of the output of the neuron. An acceptable range of output is usually between 0 and 1, or -1 and 1. Mathematically this process is described in the figure

Figure 3.11 Mathematical Model of Neurons

From this model the interval activity of the neuron can be shown to be:

$$v_k = \sum_{j=1}^{p} w_{kj} x_j$$
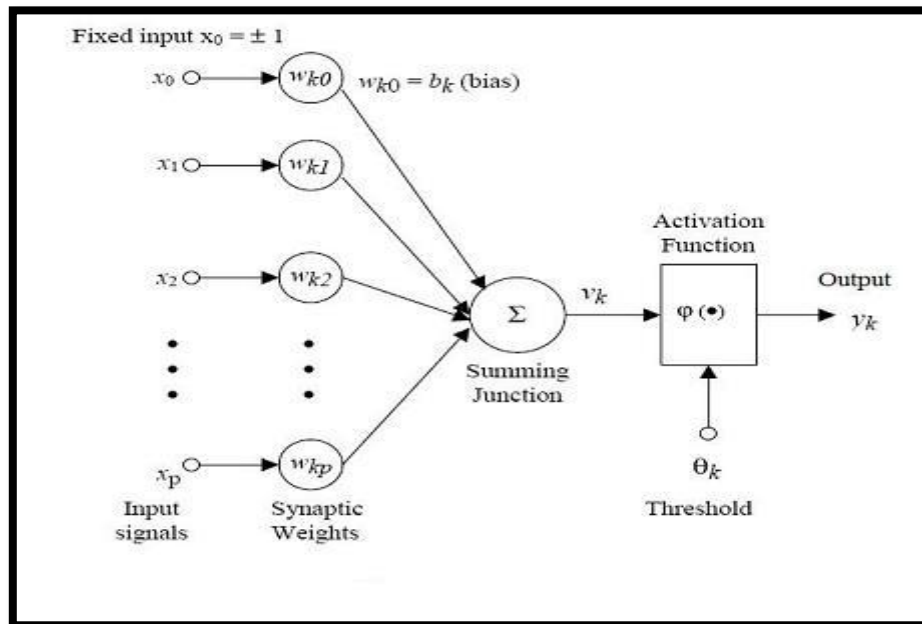
The output of the neuron, y_k, would therefore be the outcome of some activation function on the value of v_k.

3.3.8 Activation functions

The activation function acts as a squashing function, such that the output of a neuron in a neural network is between certain values (usually 0 and 1, or -1 and 1). In general, there are three types of activation functions, denoted by $\Phi(.)$. First, there is the Threshold Function which takes on a value of 0 if the summed input is less than a certain threshold value (v), and the value 1 if the summed input is greater than or equal to the threshold value.

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases}$$

Secondly, there is the Piecewise-Linear function. This function again can take on the values of 0 or 1, but can also take on values between that depending on the amplification factor in a certain region of linear operation.

$$\varphi(v) = \begin{cases} 1 & v \geq \frac{1}{2} \\ v & -\frac{1}{2} > v > \frac{1}{2} \\ 0 & v \leq -\frac{1}{2} \end{cases}$$

Thirdly, there is the sigmoid function. This function can range between 0 and 1, but it is also sometimes useful to use the -1 to 1 range. An example of the sigmoid function is the hyperbolic tangent function

$$\varphi(v) = \tanh\left(\frac{v}{2}\right) = \frac{1 - \exp(-v)}{1 + \exp(-v)}$$
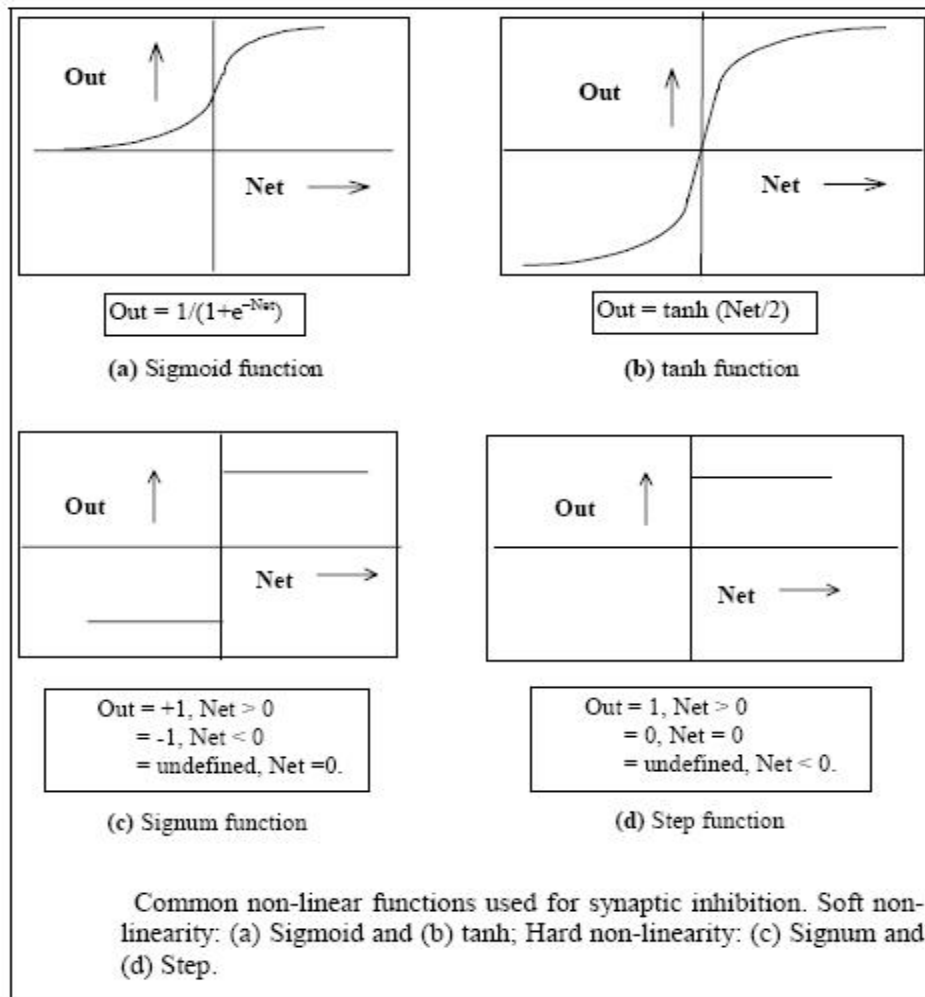
Figure 3.12 Non-Linear Functions used for Synaptic Inhibition.

The artificial neural networks which we describe are all variations on the parallel distributed processing (PDP) idea. The architecture of each neural network is based on very similar building blocks which perform the processing.

3.3.9 Neural Network Topologies

Feed-forward neural networks, where the data flow from input to output units is strictly feed forward. The data processing can extend over multiple (layers of) units, but no feedback connections are present, that is, connections extending from outputs of units to inputs of units in the same layer or previous layers.

- **Recurrent neural networks** that do contain feedback connections, contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process such

35

that the neural network will evolve to a stable state in which these activations do not change anymore. In other applications, the change of the activation values of the output neurons is significant, such that the dynamical behavior constitutes the output of the neural network.

3.3.10 Training of artificial neural networks

- **Supervised learning** or Associative learning in which the network is trained by providing it with input and matching output patterns. These input-output pairs can be provided by an external teacher, or by the system which contains the neural network (self-supervised).
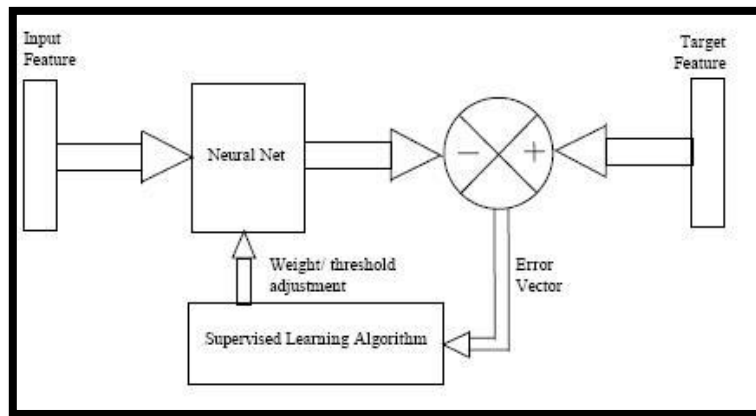
Figure 3.13 Supervised Learning Algorithm

- **Unsupervised learning** or Self-organization in which an (output) unit is trained to respond to clusters of pattern within the input. In this paradigm the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather the system must develop its own representation of the input stimuli.

Chapter 4

Proposed CADx System

4.1 Introduction

Computer-aided detection (CADx) has been developing fast in the last two decades. The main idea of CADx is to assist radiologists in interpreting medical images by using dedicated computer systems to provide 'second opinions'. The final medical decision is made by the radiologists. Studies on CADx systems and technology show that CADx can help to improve diagnostic accuracy of radiologists, lighten the burden of increasing workload, reduce cancer missed due to fatigue, overlooked or data overloaded and improve inter- and intra-reader variability [2, 5]. Recently, various types of brain Computer-Aided Detection methods have been developed by a number of researchers using brain MR images based on several types of machine learning classifiers. In general, there are two types of CADx systems for brain evaluation (i.e., systems that detect lesions and those that differentiate diseases).

Brain CADx systems can provide radiologists with a 'second opinion' to assist them in the detection of brain diseases. Consequently, radiologists expect that CADx systems can improve their diagnostic abilities based on synergistic effects between the radiologist and the computer with medical image analysis and machine learning techniques. Therefore, the CADx systems should have abilities similar to the radiologists in terms of learning and recognition of brain diseases. For this reason, pattern recognition techniques including machine learning play important roles in the development of CADx systems [3]. Pattern recognition is the act of extracting features from objects (e.g. lesions) in raw data and making a decision based on a classifier output, such as classifying each object into one of the possible categories of various patterns.

The architecture for the CADx Brain MRI system is shown in Figure 4.1. It comprises three main processes for Feature Extraction , Feature Selection, and Classification of the image.

In our research we implement a hybrid technique for medical image diagnosis system for classification of brain MRI into normal and abnormal case through the following three phases as shown in figure 4.1:

- Extracting features via Fast Discrete Curvelet Transform (FDCT)

- Selecting optimal features using Principal Component Analysis (PCA)

- Classification using Feed-Forward Artificial Neural Network with Back-Propagation (FF-ANN) algorithm.
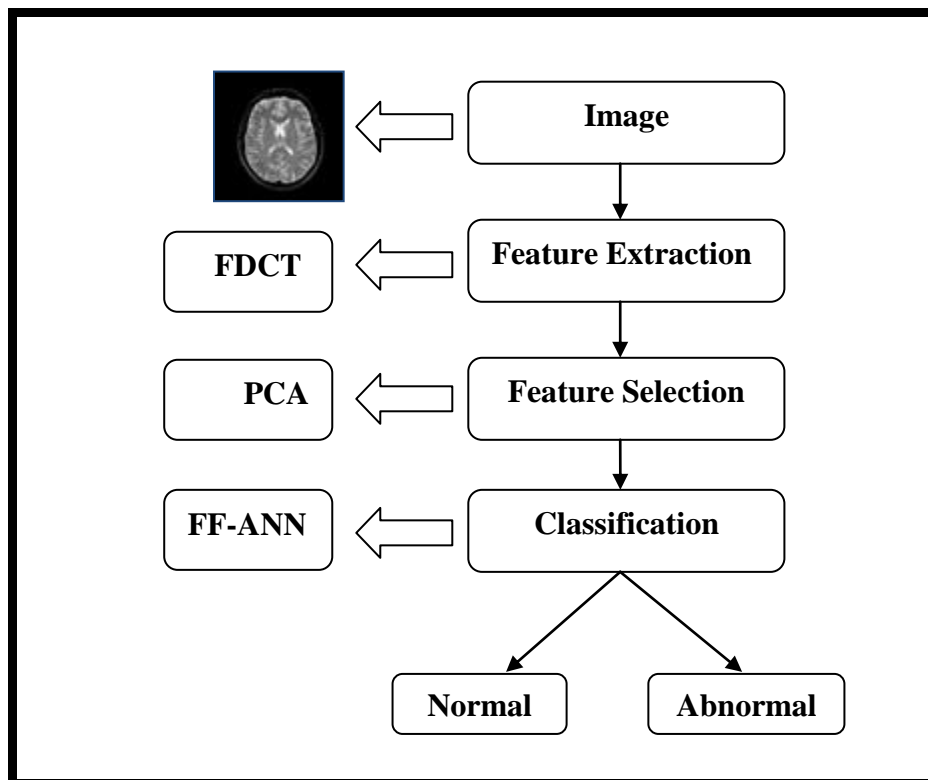
Figure 4.1. The proposed methodology of CADx system.

4.2 Feature extraction

Texture features are important in medical image diagnosis systems due to their ability to define the entire image characteristics effectively. We have already described the concepts of curvelet transform and curvelet structures at different resolutions in Chapter 3. Since discrete curvelet transform has been found to represent the curved edges of images more effectively than wavelet and Gabor filters, it is expected to find the most representative texture patterns of an image as well. In

this chapter, we describe image representation using wrapping based discrete curvelet texture features and the use of these features in CADx system . For this purpose, first, we describe the general procedure of curvelet texture features descriptor generation from spectral domain coefficients. Second, we provide the detail information on how the curvelet texture descriptors are used to represent the images in the feature database we use Finally, we provide the implementation of curvelet texture features in our CADx system. The feature extraction process was carried out through two steps: firstly the Curvelet coefficients were extracted by the DCT and then the essential coefficients have been selected by the principal component analysis PCA. In our research fast Discrete Curvelet transform via warpping will be used to extract texture features from MRI brain images.

4.2.1 Curvelet Computation

Discrete curvelet transform is applied to an image to obtain its coefficients. These coefficients are then used to form the texture descriptor of that image. Recalling "Equation 3.3" of Chapter 3, curvelet coefficients of a 2-D Cartesian grid f [m,n] , $0 \leq m < M, 0 \leq n < N$ are expressed as:

$$C^D(j, l, k) = \sum_{\substack{0 \leq m < M, \\ 0 \leq n < N}} f[m.n] \, \overline{\varphi^D_{J,l,k}}[m,n]$$

where $\overline{\varphi^D_{J,l,k}}[m, n]$ is the curvelet mother function. This transform generates an array of curvelet coefficients indexed by their scale j , orientation l and location parameters (k_1 , k_2) . The wrapping based curvelet transform is calculated as follows [18] :

- Apply the 2D FFT and obtain Fourier samples $\hat{f}[m, n]$ of the image.
- Find the product of Fourier domain digital curvelet waveform and the Fourier samples of the image.
- Wrap the product around the origin to obtain rectangular support.
- Apply the inverse 2D FFT on the wrapped product.

Multiresolution discrete curvelet transform in the spectral domain utilizes the advantages of fast Fourier transform (FFT). During FFT, both the image and the curvelet at a given scale and orientation are transformed into the Fourier domain. The convolution of the curvelet with the image in the spatial domain then becomes their product in the Fourier domain. At the end of this computation process, we obtain a set of curvelet coefficients by applying inverse FFT to the spectral product.

This set contains curvelet coefficients in ascending order of the scales and orientations.

The complete feature extraction process using one single curvelet is illustrated in Fig. 4.2(a). There is a problem in applying inverse FFT on the obtained frequency spectrum. The frequency response of a curvelet is a trapezoidal wedge which needs to be wrapped into a rectangular support to perform the inverse Fourier transform. The wrapping of this trapezoidal wedge is done by periodically tiling the spectrum inside the wedge and then collecting the rectangular coefficient area in the origin. Through this periodic tiling, the rectangular region collects the wedge's corresponding fragmented portions from the surrounding parallelograms. For this wedge wrapping process, this approach of curvelet transform is known as the 'wrapping based curvelet transform'. The wrapping is shown in Fig. 4.2(b) in

order to do IFFT on the FT wedge, the wedge has to be arranged as a rectangle. The idea is to replicate the wedge on a 2-D grid, so a rectangle in the center captures all the components a, b, and c of the wedge.

Wedge wrapping is done for all the wedges at each scale in the frequency domain, so we obtain a set of subbands or wedges at each curvelet decomposition level. These subbands are the collection of discrete curvelet coefficients.

- **To provide an illustration of curvelet subbands**,

In order to extract feature vector, the image is decomposed into its approximate and detailed components using three levels of Curvelet transform. These sub-images thus obtained are called image *Curvelet subbands*. Which greatly reduces the dimensionality of the original image. Thereafter only the approximate components are selected to perform further computations, as they account for maximum variance. Thus, a representative and efficient feature set is produced.
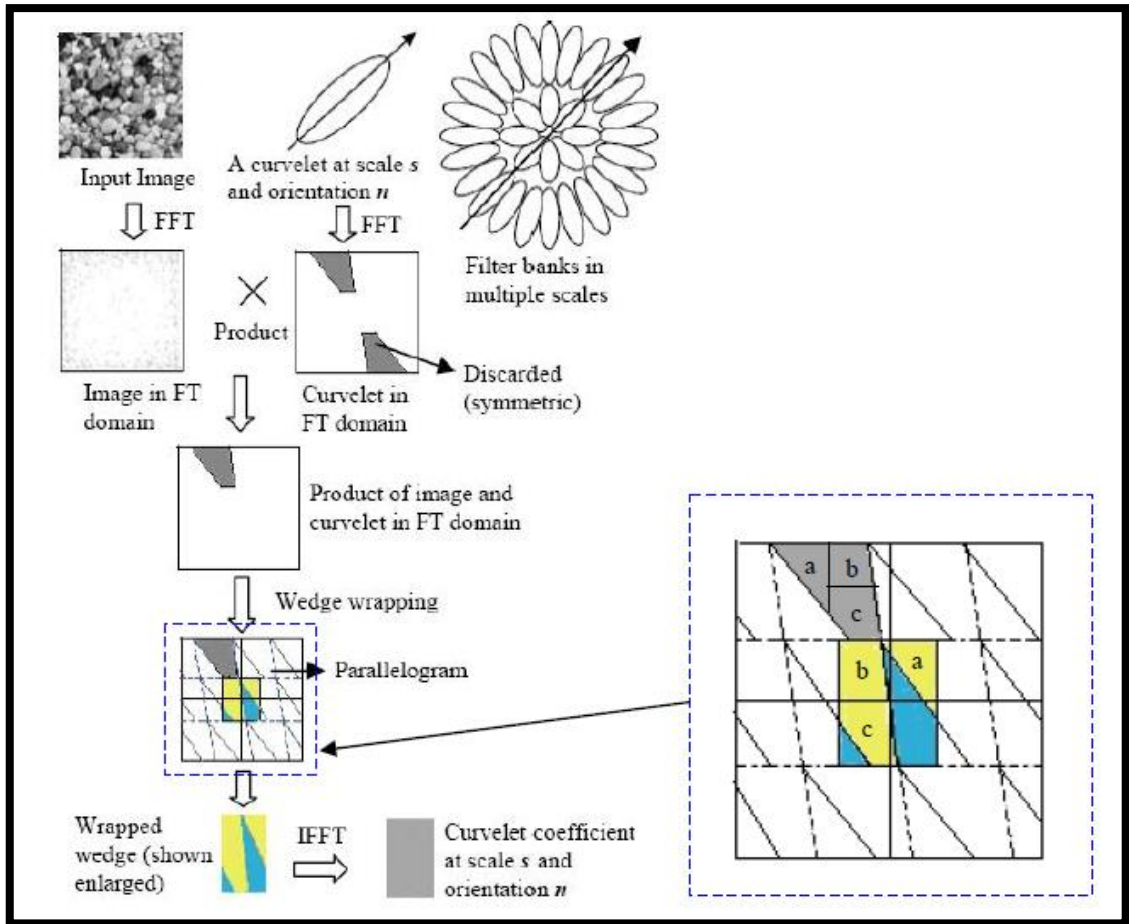
Figure. 4.2: Fast discrete curvelet transform to generate curvelet coefficients.

In more details, given an image I with size = [M,N], we compute the number of scales as log2(min(M,N)) - 3)] and For each scale of them, a different number of angles(must be a multiple of 4) is applied .

We apply fast discrete curvelet transform to 256×256 MRI Brian image shown in figure 4.3(a) using Curvelab-2.1.2 of [46]. Figure 4.3(b) shows the Curvelet coefficients of this Image decomposed at scale = 5 and orientation [1,8,16,16,1] for each scale respectively. The coarse scale (low frequency) is the image in the center, which so-called approximate coefficients. The other images are finest scale which so-called detail coefficients.
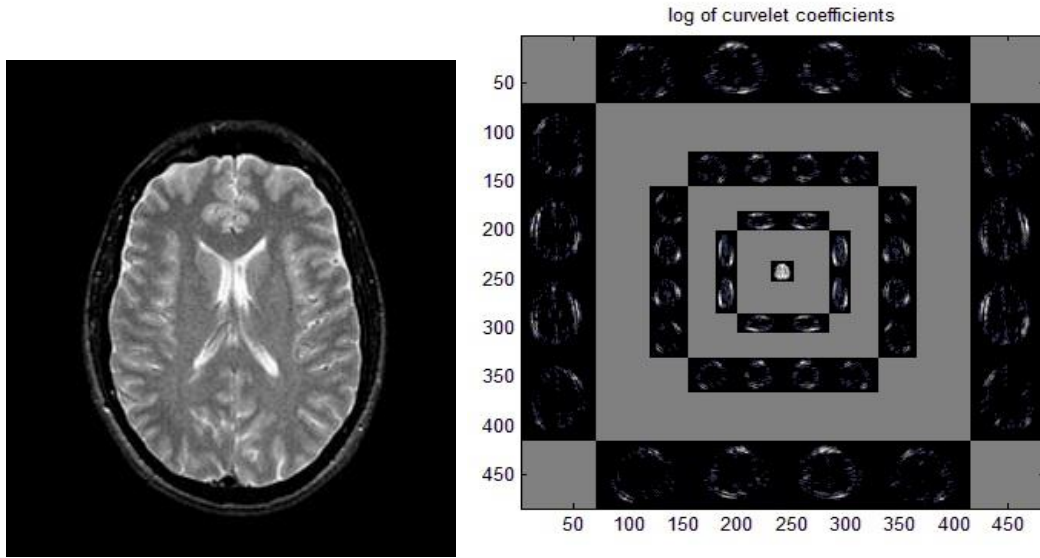
Figure 4.3: Example of Curvelet coefficients: Original image (left), Coefficients (right, scale=5).

The Curvelet coefficients in figure 4.3(left) are described as follows:

- The low frequency (coarse scale) coefficients are stored at the center of the display.
- The Cartesian concentric coronae show the coefficients at different scales; the outer coronae correspond to higher frequencies.
- There are four strips associated to each corona, corresponding to the four cardinal points; these are further subdivided in angular panels.
- Each panel represent coefficients at a specified scale and along the orientation suggested by the position of the panel.

4.2.2 Features Extraction For the Training Database using Curvelet

Our first step is to perform the wrapping based discrete curvelet transform using Curvelab-2.1.2 toolbox from [46] to find the coefficients which capture image's features at different angles and different scales. In order to create feature vectors for every 256×256 images in the database. five levels of decomposition is used which is the highest level of decomposition possible for a 256×256 image using Curvelab-2.1.2. **In this implementation**, We obtain one subband at the coarsest and one subband at the finest level of curvelet decomposition. For other levels of decomposition, we get different number of directional subbands at each level which

is shown in Table 4.1. Because of the symmetry of the directional bandpass images, at each scale only one half of the directional bandpass images are sufficient for feature extraction. The reason why only first half of the total subbands at a resolution level are considered for feature calculation is that the curvelet at angle θ produces the same coefficients as the curvelet at angle (θ+π) in the frequency domain. These subbands are symmetric in nature. Therefore, considering half of the total number of subbands at each scale reduces the total computation time for the feature vector formation.

Table 4.1: Curvelet subband Distribution at Each Scale

Curvelet transform					
Scale	1	2	3	4	5
Total no. subbands	1	8	16	16	1
Subband considered for feature calculation	1	4	8	8	1

A given brain image of size 256×256 is decomposed into 5 scales, 8 directions at scale 2, 16 directions at scale 3, and 16 directions at scale 4, then, after excluding redundant directional subband images, there are 20 directional bandpass images for feature extraction. The size of the directional bandpass image varies from scale to scale, for example, at scale 2 there are 8 bandpass images with size 19×43 Similarly at scale 3, eight bandpass images of size 35×44, 8 images of size 32×42 and at scale 4, eight bandpass images of size 70×68, 8 images of size 64×86.

Once the curvelet coefficients are generated for all subbands, The magnitudes of the significant coefficients and their positions have been concatenated and used as a texture feature vector of the image called curvelet feature descriptor (CFD)that used to represent images in the feature database.

for each of the 256×256 images in the database when using 5 levels of discrete curvelet decomposition We obtain a total of $(1+4+8+8+1) = 22$ subbands. 20 directional subbands and the coarsest subband at first scale and the finest subband at the last scale. After the feature descriptors are generated. All the images in the image dataset are represented by their curvelet feature descriptor (CFD) in the feature database. For our database, the feature database size is N×CFD where N is the no. of

the database images and CVD is the curvelet feature descriptor. The Pseudocode for Texture Feature Extraction using FDCT is described in table 4.2 below.

Table (4.2): Pseudocode for Texture Feature Extraction using FDCT

Purpose: Extract texture features descriptor from an image.

Input: 256x256 brain images.

Output: Curvelet feature descriptor (CFD).

Procedure:

{

For each image in Dataset

  {

  Step1: Read the input image.

  Step2: Apply FDCT on the image.

  Step3: Calculate curvelet coefficients of MRI image for each scale with different orientations.

  Step4: Concatenate the curvelet feature coefficients of all regarded subbands.

  Step5: Return the texture vector descriptor generated from step 4.

  }

}

4.2.3 Feature Selection via PCA

The number of the obtained Curvelet coefficients for each image is very high. So, according to evidences, performing all Curvelet coefficients to classifiers is not suitable and may degrade the system performance .

As soon as we generate the feature vector descriptor from feature extraction stage, we either feed it further as input to the classification algorithm, or employ a feature selection algorithm in order to omit redundant or irrelevant features and maintain the most informative and discriminatory ones, facilitating at the same time the classification process

In this research, a well-known dimensionality reduction tool, PCA, is applied .

 Only important Curvelet subbands coefficient are selected depending on the amount of total variance they account for. The main idea behind using PCA in our approach is to reduce the dimensionality of the curvelet feature descriptor.

PCA is applied to a matrix X containing *N* training dataset images, each *Xi* is a row vector which represent curvelet feature descriptor (CFD) for a sample image *i,* where *i*=(1,..,n). Then we can get a lower dimensional representation of the coefficients for classification. This is not only reduces computational load, but also increases recognition accuracy.

The input feature space is transformed into a lower-dimensional feature space using the largest eigenvectors of the correlation matrix such that the variance of the reconstructed data is preserved. PCA provides optimal solution with a low computational cost and computational complexity [13]. The Algorithm that describes PCA method is listed in table 4.3 below.

Table (4.3): PCA Pseudocode

Purpose: To extract optimal texture features from feature vector descriptor.

Input: Curvelet feature descriptor (CFD) of length M, N: # of images in dataset.

Output: Curvelet feature descriptor (CFD$_1$) of length L where L<<M.

Let X be an input dataset (X: matrix of dimensions M×N).

Procedure:

{ **Step 1. Calculate the empirical mean:**

$$u[m] = \frac{1}{N} \sum_{n=1}^{N} X[m, n].$$

Step 2. Calculate the deviations from the mean and store the data in
the matrix $B[M \times N]$: $B = X - u.h$, where h is a 1 × N row vector of all 1's:
h[n] = 1 for n = 1, . . . , N.

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

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

Step 5. Rearrange the eigenvectors and eigenvalues: $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N$

Step 6. Choosing components and forming a feature vector: save the first L columns of V as the M × L

matrix W, $W[p, q] = V[p, q]$, $for\ p = 1, \ldots, M, q = 1, \ldots, L\ where\ 1 \ll L \ll M$

Step 7. Deriving the new dataset: The eigenvectors with the highest eigenvalues are projecte

space,

 this projection results in a vector represented by fewer dimension (L < M) containing the esse

coefficients.

Step 8. Return the essential coefficients vector (CFD$_1$).

}

Now the result of this stage curvelet feature descriptor CVD$_1$ consists of only the most discriminative principal components for each image in dataset. The size of the input CVD is reduced from (M) to (L) where L<<M.

The feature extraction process was carried out through two steps: firstly the curvelet coefficients were extracted by the FDCT and then the essential coefficients have been selected by the PCA.

4.3 Classification using artificial neural network (ANN)

ANN is a mathematical model consisting of a number of highly interconnected processing elements organized into layers and functionality of which have been resembled to that of the human brain [17]. ANN have been widely used in many pattern recognition problems, such as medical image classification, optical character recognition, and object recognition. Since brain image classification can be treated as a two class pattern recognition problem. The advantage of using neural networks for classification is the feasibility of training a system.

The architecture of artificial neural network used in this thesis is a multi layer feed forward network with steepest descent back-propagation training algorithm with adaptive learning rate. Figure 4.4 shows the architecture of a typical Back-propagation neural network. It contains from three layers; input layer, one hidden layer, and output layer. Log sigmoid transfer function was used in the hidden layer. Linear transfer function was used in the output layer. The parameters of the network were adjusted by training the network on a set of training images dataset. The training of the network was performed under back propagation of the error, by using minimum square error (MSE) as objective function
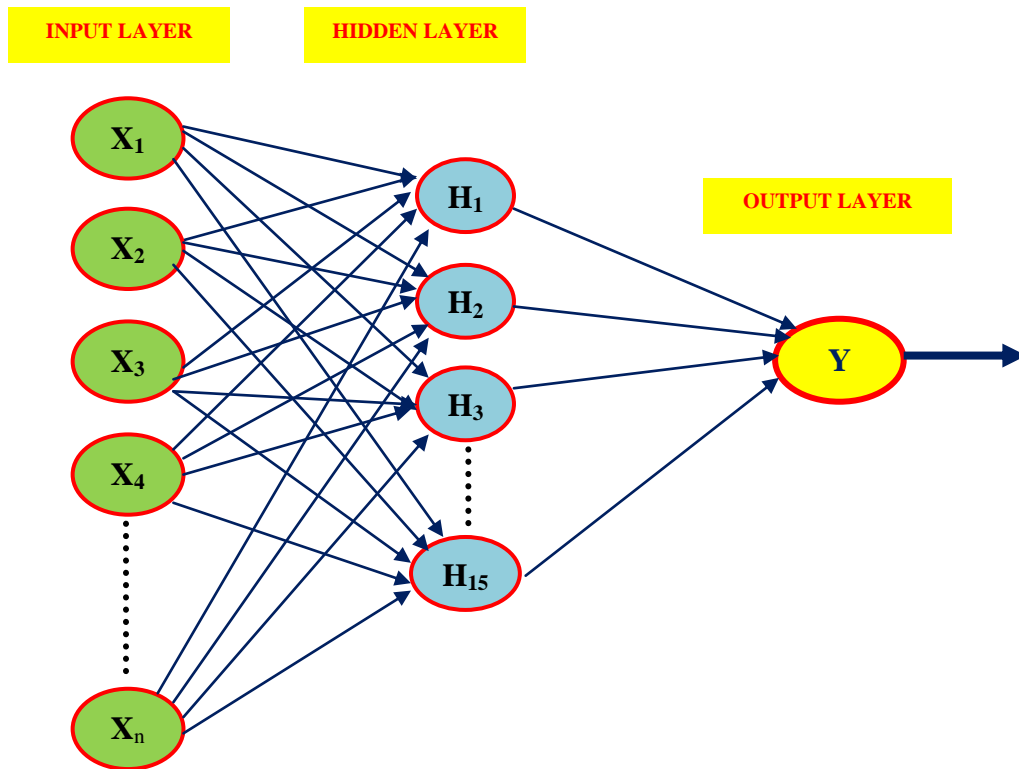
Figure 4.4: Back-propagation Neural Network.

The trained networks were then be used to predict labels of the new data.

The gradient descent with momentum training function was used to update weight and bias values. The number of hidden layer and the number of nodes in each layer are varied to obtain the optimal back-propagation neural network topology for highest classification performance.

As known, the number of hidden layer and the number of nodes in each one are important parameter to improve the recognition rate. The first layer consisted of input elements in accordance with the feature vectors selected from the curvelet feature descriptor by the PCA. The number of neurons in the hidden layer was fifteen. One neuron in the output layer was used to represent normal or abnormal class. The details of back-propagation (BP) algorithm are well documented in the (chapter 3 section 3.3).

The neural network has been trained to adjust the connection weights and biases in order to produce the desired mapping. At the training stage, the feature vectors are applied as an input to the network and the network adjusts its variable parameters,

the weights and biases, to capture the relationship between the input patterns and outputs .The performance is measured by mean squared error (MSE):

where *Yex* is the target value, Ymod is the actual output, and n is the number of training data.

Table (4.4): ANN Pseudocode with Back Propagation
*Purpose***:** Classify brain images.

*Input***:**   Curvelet feature descriptor (CVD1) for each image in training dataset.

*Output***:**   Normal or abnormal brain.

Procedure: The back-propagation algorithm includes the following steps:

{
1. Initialize weights and biases to small random numbers.
2. Present a training data to neural network and calculate the output by propagating the input forward through the network used.
3. Propagate the sensitivities backward through the network:
4. Calculate weight and bias updates
5. Input the features descriptor of test image on the trained neural network.
6. Classify image into normal and abnormal
}

4.4. Proposed System

The proposed system consists of two phases as shown in the block diagram of Figure 4.5: an offline phase and an online phase. Both the phases, consist of the following steps: feature extraction based on FDCT , feature reduction through PCA and classification by FF-ANN classifier.

- **Offline Phase**:

Let n be the number of training images . The steps of the offline (training) phase are listed as follows:

Step 1: The training n images are decomposed by FDCT to get the feature vector (Curvelet feature descriptor).

Step 2: The dimension of the feature vectors representing the training images is reduced by applying PCA as mentioned in (section 4.2.3)

Step 3: The set of reduced feature vectors, along with the class information are used to train FF-ANN classifier.

- **Online Phase:**

The online phase of the proposed system consists of the following steps:

- Step 1: The user (doctors, radiologist, etc.) inputs the brain MR image to be classified. FDCT is applied on the input image to get the feature vector (Curvelet Coefficients)

- Step 2: The dimension of the feature vector representing the input test image is reduced by applying PCA.

- Step 3: This reduced feature vector is used as input to the previously trained FF-ANN classifier. The classifier classified the input image as normal or abnormal.
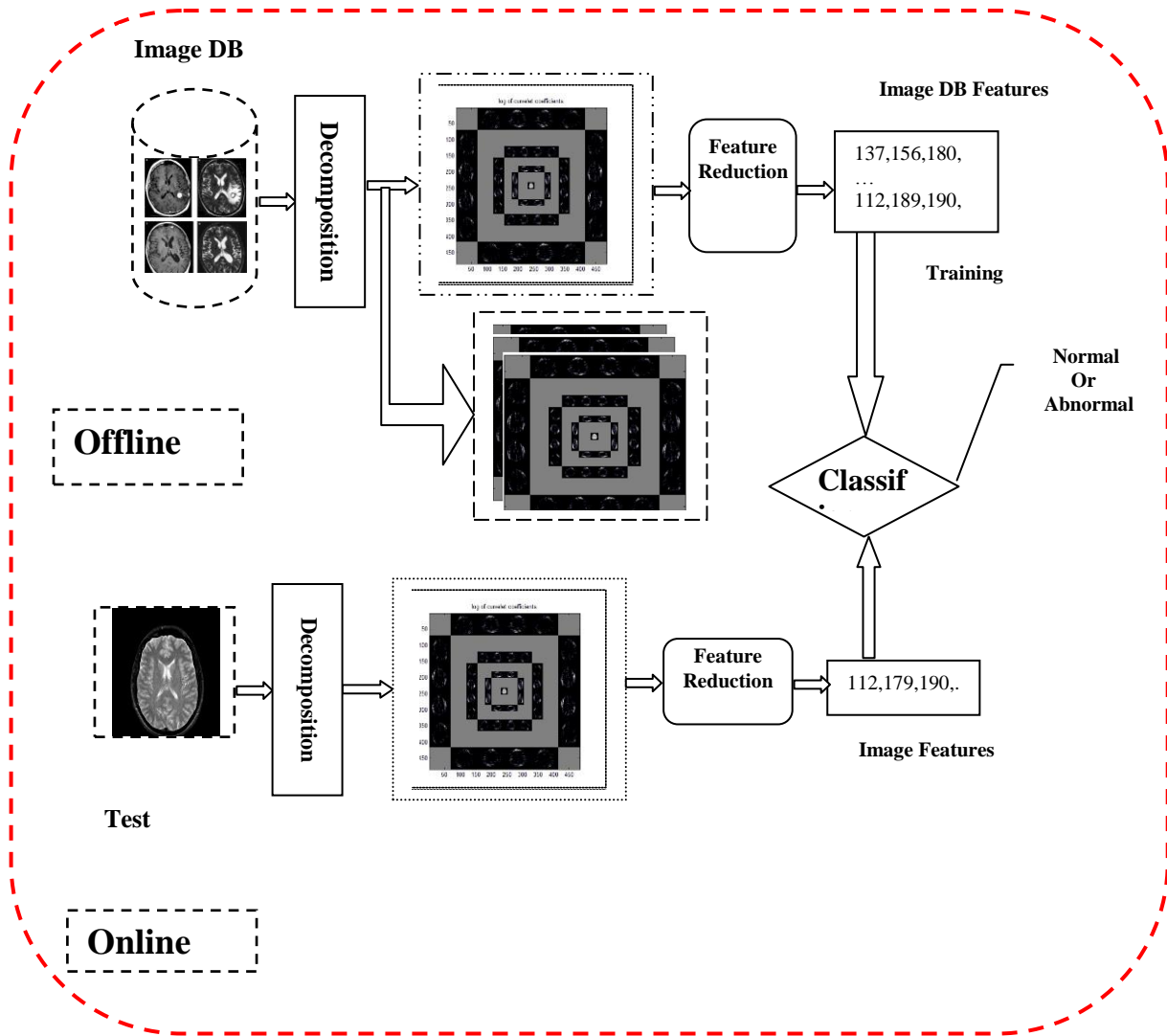
Figure 4.5. Block diagram of the overall proposed system.

Chapter 5

Experiments and Results

5.1 Introduction

In previous chapters, we described the concept of curvelet, its implementation and feature extraction process. Based on those analyses, we find that curvelet is powerful and effective in capturing edge information accurately which is a key to texture features based CADx. In this chapter, we test the performance of medical image diagnosis system using curvelet texture features. We also test curvelet features performance on the scale distorted image diagnosis.

In the literature, wavelet texture feature is the effective bench mark for MRI brain image classification. In order to benchmark the wrapping based curvelet texture diagnosis performance, we compare its performance with that of existing wavelet texture features.

For this purpose, we organize our experiments in the following ways:

1. First, we need to find the optimal curvelet decomposition level that provides the best texture feature descriptor for improving CADx classification performance. From the curvelet texture representation shown in Fig. 4.3 of Chapter 4, we find that image texture pattern is clearer at the higher levels (4, 5 and 6) than at the lower levels (1, 2 and 3) when we use Curvelab-2.1.2 [46]. In chapter 4, we created curvelet texture descriptors using 5 levels of decomposition. Now we compare MRI brain image classification performance of these curvelet features to determine the best performance of curvelet.

2. Finding the robustness of discrete curvelet texture features to scale distortions by performing MRI brain image classification experiments on a large database that contains original and scale distorted images.

3. To benchmark the best performance of curvelet texture feature, we compare the curvelet classification performance with that of the wavelet and other methods.

4. To measure the overhead of our proposed system, we calculate the computational time needed and compare it with some of the other methods used in the literature.

5.2 **MRI Database**

The dataset used in this thesis consists of T2-weighted MR brain images in axial plane and were downloaded from the website of Harvard Medical School (URL: http://med.harvard.edu/AANLIB/) [32]. The benchmark dataset (Dataset-160) consists of 160 (20 normal and 140 abnormal) brain MR images. Three different resolutions of Dataset-160 were used in the experiments: 256×256, 128×128 and 64×64 pixels. So we have 3 different datasets. 128 test images are used to train the neural network and classified by the proposed hybrid techniques into normal or abnormal brain. 32 images are used to test the system which are described in detail in Table 5-1. Fig. 5.1, shows samples of the brain MR images used in the experiments.

Table 5.1: The experiments dataset.

Dataset (160)		Training (128)		Testing (32)	
Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
20	140	16	112	4	28
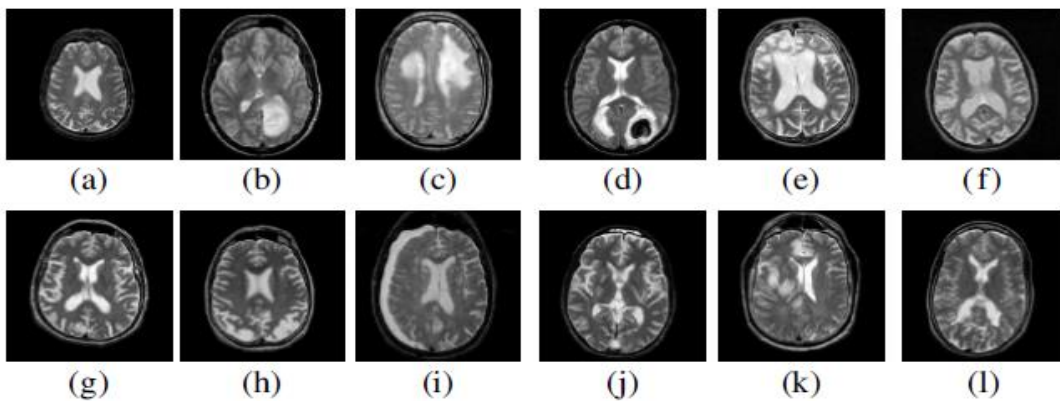
Figure 5.1 Sample brain MR images: (a) Normal, (b) glioma, (c) meningioma, (d) sarcoma, (e) Pick's disease, (f) Huntington's disease, (g) Alzheimer's disease, (h) Alzheimer's disease with visual agnosia, (i) chronic subdural hematoma, (j) cerebral toxoplasmosis, (k) herpes encephalitis, (l) multiple sclerosis.

5.3 Implementation Environment

Proposed CADx system is implemented using MATLAB program of version 7.8.0.347 (R2011a) which imbedded with image processing tools and using

Curvelab-2.1.2. We performed all the computations of FDCT + PCA + FF-ANN classification on a personal computer with a platform of Intel Core i3 Processing power of 2 GHz CPU with 4 GB RAM during the implementation. 160 image dataset went through our implemented system to extract the texture features and store them. The extracted features are reduced to the most representative features by using PCA and they are used for classification using the FF-ANN with back propagation algorithm. This step is made offline for all images in the database. Now the database and ANN are ready for testing and evaluating our CADx proposed system.

5.4 Performance Evaluation Metrics of CADx Systems

In this section, we present the performance evaluation methods used to evaluate the proposed approaches. We evaluate the performance of the proposed method in terms of sensitivity, specificity , accuracy. These terms are defined as follows [19]:

- **Sensitivity** (true positive fraction) is the probability that a diagnostic test is positive, given that the person has the disease,

$$sensitivity = \frac{TP}{TP + FN}$$

- **Specificity** (true negative fraction) is the probability that a diagnostic test is negative, given that the person does not have the disease,

$$specificity = \frac{TN}{TN + FP}$$

- **Accuracy** is the probability that a diagnostic test is correctly performed,

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

where: TP (True Positives) – correctly classified positive cases,

TN (True Negative) – correctly classified negative cases,

FP (False Positives) – incorrectly classified negative cases, and

FN (False Negative) – incorrectly classified positive cases.

5.5 Proposed System Evaluation

Here, we come to the most important section in the thesis. In this section, we test our proposed system and show the results. We explain the results, comment on them, and compare our system with other existing systems. We evaluate the system regarding three metrics: Sensitivity, Specificity, and Accuracy.

5.5.1 Sensitivity, Specificity, and Accuracy Evaluation

In order to assess the efficiency of the proposed technique described in the previous chapter, a series of experiments were carried out using all datasets separately. Therefore, for each 256x256, 128x128 and 64x64 dataset, 128 images were used for training and 32 images for testing. In this experiment and for each dataset, training images were submitted to proposed system. These images are decomposed using curvelet transform at different scales and orientations. For example, a given brain image of size 256×256 is decomposed into 5 scales, 8 directions at scale 2, 8 directions at scale 3, and 8 directions at scale 4, therefore 26 (1+8+8+8+1) subbands are produced, including 2 approximate (finest and coarsest) and 24 detailed subbands. Then, after excluding redundant directional subband images, there are 12 directional bandpass images for feature extraction. The size of the directional bandpass image varies from scale to scale, for example, at scale 2 there are 8 bandpass images with size 19×43. At scale 3, 8 bandpass images of size 38×85, and 8 sub images of size 75×171 as scale 4. To further reduce the dimensionality and feature selection, PCA is applied on these details components only. After curvelet sub-images are projected to eigenspace, the first 60 principal components were selected to reduce the feature vector size to only 60 (12*5), where these features, preserving the variance of the original decomposed image features. neural network classifier for MRI normal/abnormal is employed to perform the classification task. In this stage the neural network is used to classify brains into normal or abnormal class. it is a 3 layers simple feed forward network. The number of neurons in the input layer is 60 with 1 neuron in the output layer and 15 neurons in the hidden layer and two different activation functions are used for the network. The sigmoid function is used between the input layer and hidden layer while the linear output function is used between the hidden and output layer. A neural network starts

by initializing the weights and bias randomly. The neural network is trained using Levenberg-Marquardt algorithm with the Mean Squared Error (MSE). Then Sensitivity, Specificity, and Accuracy were calculated. The results are shown in Table 5.2.

Table 5.2: Sensitivity, specificity and accuracy values for the classification of MR images as normal and abnormal via curvelet transform with PCA.

Image Size	No. of scales	TP	TN	FP	FN	Sensitivity %	Specificity %	Accuracy %
256x256	5	27	4	0	1	96.4	100	96.9
128x128	4	27	4	0	1	96.4	100	96.9
64x64	3	26	4	0	2	92.8	100	93.7

Table 5-2 shows the classification rates in terms of Sensitivity, Specificity and Accuracy for the proposed approach. PCA is used for feature reduction for reducing the computation complexity of the system, which was described in Section 3.2

The dimension of the feature vector was reduced to 60 features with the PCA algorithm. Limiting the feature vectors to the component selected by the PCA leads to an increase in accuracy rates. In this experimental, MRI dataset that have healthy and diseased brain are classified by the proposed classifiers. The analysis of the experimental results shows that classification accuracy 96.9% is achieved with the FF-ANN classifier
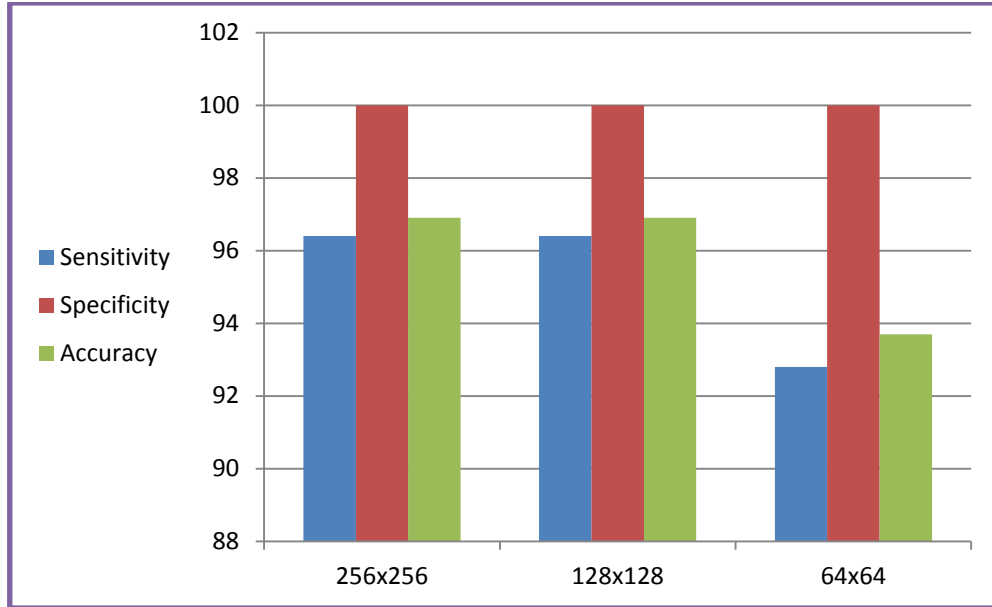
Figure 5.2: Sensitivity, Specificity, and Accuracy of different dataset resolution.

5.5.2 Time complexity

In addition to Sensitivity, Specificity, and Accuracy, an important factor in CADx system is running time for whole system stages. For Feature Extraction process, the average time that our proposed system takes for feature extraction per one image using fast discrete curvelet transform with warping is shown in table 5.3 below.

Table 5.3: Average Running time.

Image Size	Feature Extraction Average Time (Sec.)	No. of decomposition scale
256x256	0.472530	5
128x128	0.173511	4
64x64	0.069054	3

Figure 5.3 shows the time comparison between our proposed system and (Wavelet+PCA). Wavelet + PCA feature extraction time was 0.301 seconds, 0.135 seconds and 0.041 for 256x256 dataset, 128x128 dataset and 64x64 respectively. This time is increased when using the Curvelet + PCA to become 0.472 seconds for 256x256 dataset, 0.174 seconds for 128x128 dataset and 0.069 for 64x64 since the high redundancy which suffers from.
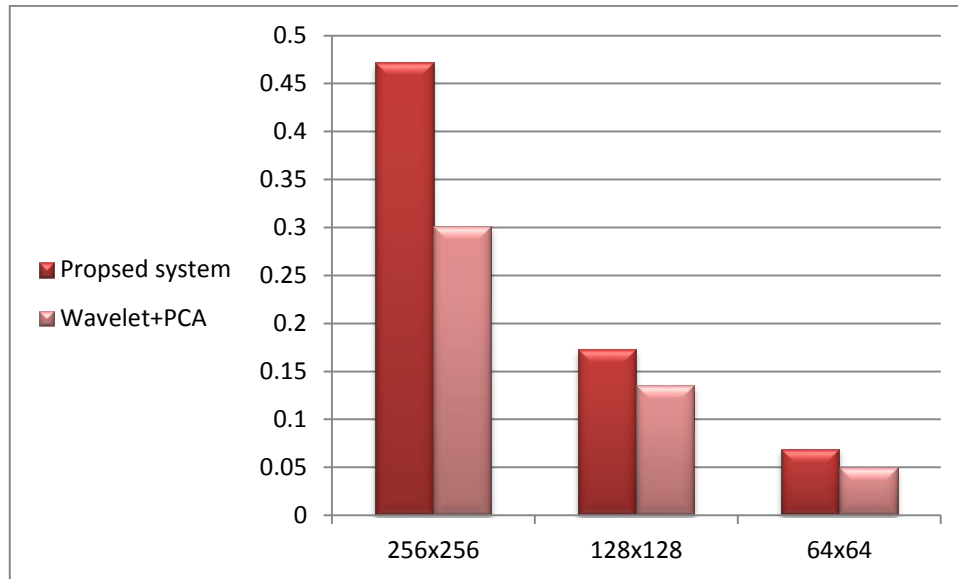
Figure 5.3: Average time for feature extraction.

5.5.3 Number of Principal Components

To find out the proper number of principal components, which give the best result, the performance of the proposed system was experimented with different numbers of principal components (10-100) for the three different datasets. The graphs of the Figures. 5.4, 5.5, and 5.6, show the performance of the proposed system in terms of Sensitivity, specificity and accuracy for the three datasets used in the experiments with different numbers of principal components, respectively. It is clear from the results given in Figs.5.3,5.4 and 5.5, that our proposed system works efficiently for all the three datasets using only 60 principal components for image representation. We have achieved the best results for all the statistical measures used to evaluate the performance of the proposed system, considering only 60 principal components with Sensitivity (96.4%, 96.4%, 92.8%), specificity (100%, 100%, 100%) and classification accuracy (96.9%,96.9%,93.7%) for three datasets respectively. In addition, These results are almost similar, which show the great capabilities of Curvelet transform feature extraction and how it does not affected by scale distortion of the low resolution dataset.
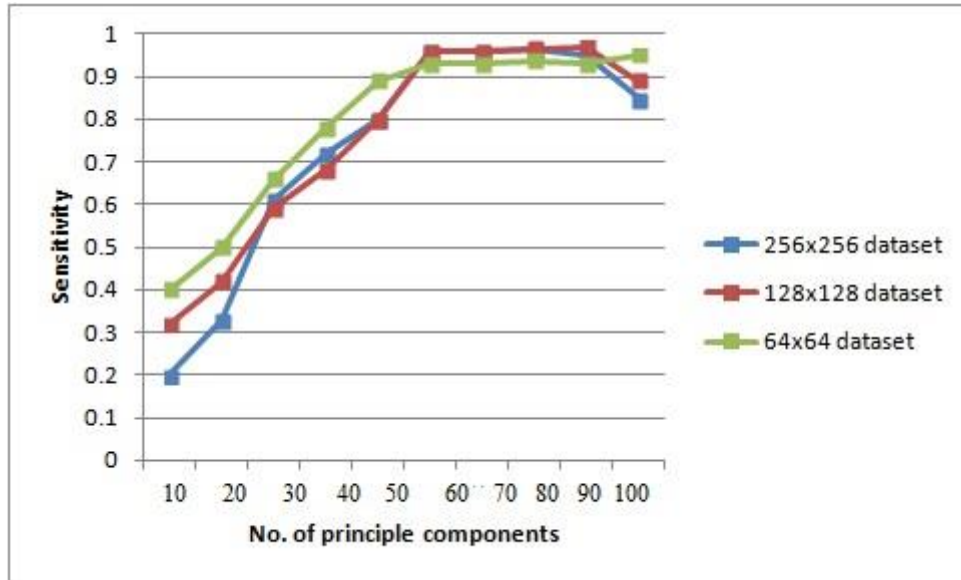
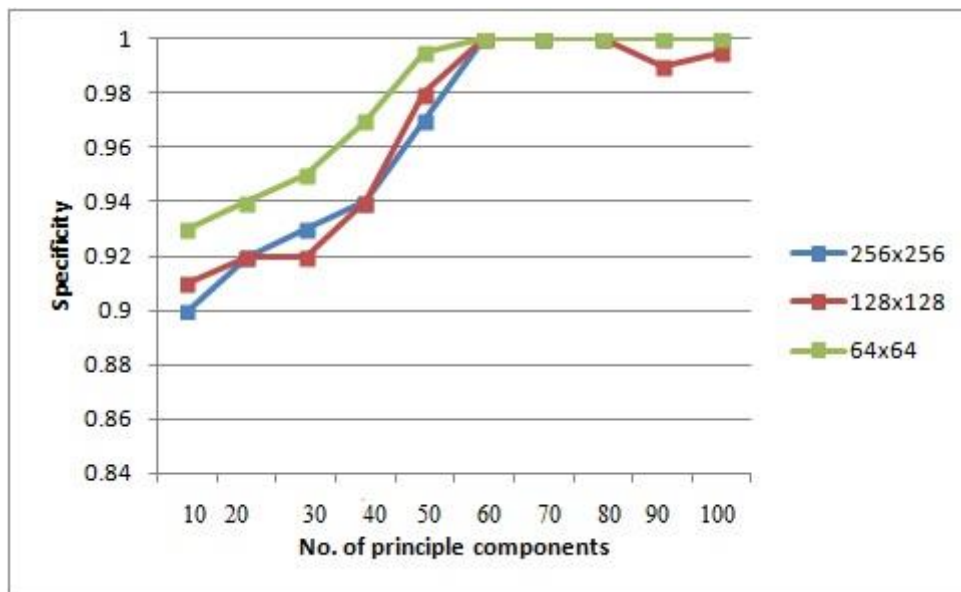Figure 5.4: Sensitivity of different dataset resolution.

Figure 5.5: Specificity of different dataset resolution.
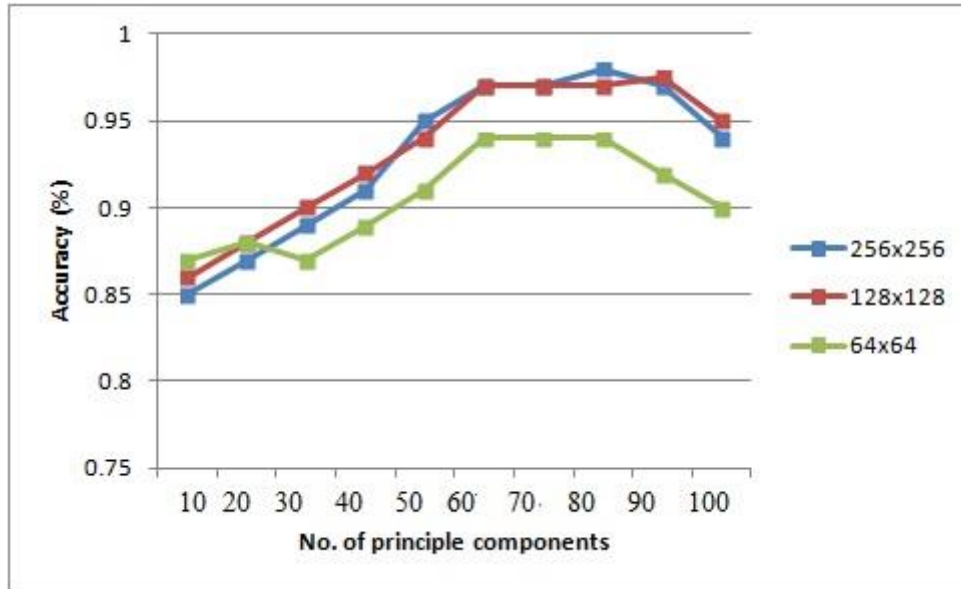
Figure 5.6: Accuracy of different dataset resolution.

5.6 Comparison of Proposed System with Other Systems

In this section, we compared the performance of the proposed system with some state-of-the-art brain MR image classification schemes.

To evaluate the effectiveness of our method we compare our results with recently results [47,49,50] for the same MRI datasets. The comparison results in table 5.4. shows the classification accuracies of our method and state of arts methods. This comparison shows that our system has higher classification accuracy.

Table 5.4: Comparison of Proposed System with Other Systems.

Scheme	Accuracy (%)
DWT + PCA + ANN	95.7
Gabor +PCA+ ANN	91.8
DWT+PCA+SOM	93.2
Proposed Method	96.9

Figure 5.7: Comparison of different classification systems.

Chapter 6

Conclusion and Future Work
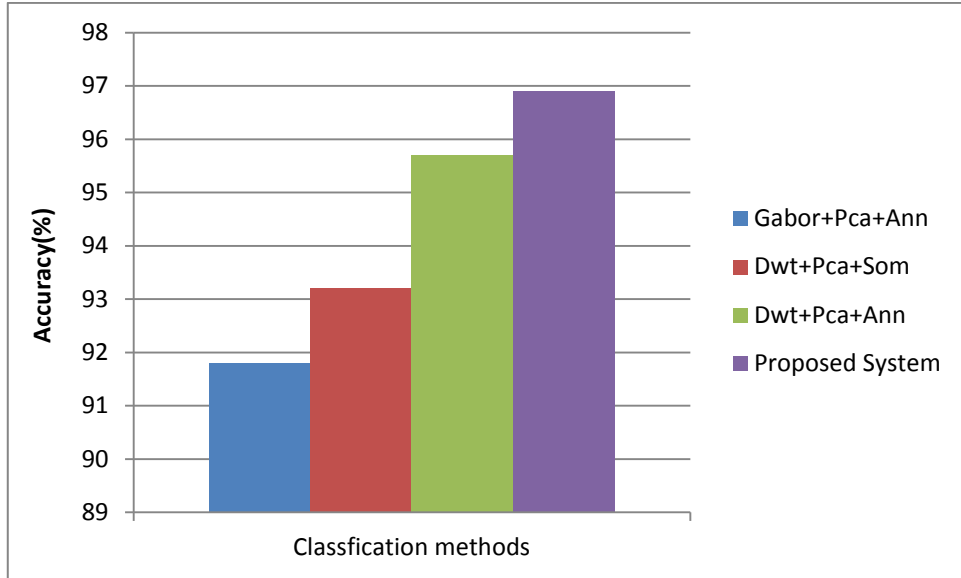
6.1 Conclusion

Cancer is the second leading cause of death for both men and women worldwide and is expected to become the leading cause of death in the next several decades. early detection and diagnosis of cancer increase the treatment options.

Magnetic Resonance Imaging (MRI) is one of the best medical Imaging technologies that currently being used for diagnosing brain tumor. The amount of medical images used to for diagnosis purposes has significantly increased due to the increased number of examinations. This increased volume of medical data makes their management difficult for medical centers, which result in an increase in the time needed to access the data, the patient's stay time in hospitals, the number of unnecessary examinations and the cost of health care.

Medical images are strongly deteriorated by noise due to data acquisition systems, various sources of interference, operator performance which can lead to serious classification inaccuracies. So it becomes very important to find new method aids in detection and diagnosis of brain tumor . The computer-aided diagnosis (CADx) is a helpful system used to classify MR brain images into normal or abnormal categories to help the radiologists and physicians.

(CADx) system consists of three stages namely feature extraction, feature reduction, and classification .The performances of different (CADx) can come to the same level of classification accuracy with the same features extracted from medical image.

To improve existing CADx performance, it is very important to find effective and efficient feature extraction mechanisms. In This research we aim to improve the performance of CADx using texture features.

Texture features is one of the most important and prominent properties of an image we have investigated Several methods that described in recent literature using both the spatial and spectral approaches. Most of them are not capable of effective texture representation .Curvelet transform, provide a good opportunity to extract more accurate texture features suitable for use in image classification . In this thesis, we

proposed curvelet transform as it provides more detailed information of the image in the spectral domain using more orientation information (multidirectional features) at each scale. Therefore, our thesis are summarized as following.

1. Based on our study we proposed a new hybrid scheme for CADx system investigate multiresolution spectral features used in CADx. Extensive studies are made to find out a suitable spectral approach of texture features representation. From the CADx background, we analyzed and examined the advantages and disadvantages of several spectral methods such as wavelet transform and Gabor filter. Therefore, we study the spectral methods used in CADx and try to investigate the reasons of the drawbacks. We found Multiresolution discrete curvelet transform is then an effective approach to represent image textures.

2. We studied the existing methods of curvelet transform. Based on this study, we found wrapping based fast discrete curvelet transform is the most efficient and effective among the existing curvelet transforms . then we studied and analyzed this method to understand its construction and its advantages in representing texture of an image. Related works using the discrete curvelet transform are also investigated.

3. We applied and evaluated discrete curvelet texture features for image classification. for CADx, it is necessary to index all the images in a large database in terms of their features. Therefore, wrapping based discrete curvelet transform is used to represent database images for classification. To evaluate the robustness of our proposed method, curvelet texture features are computed for different dataset image resolutions. CADx tests are performed using database containing 160 images obtained from Faculty of medicine in Harvard university. Experimental results for the proposed method achieved an overall classification accuracy of 96.9%, with 96.4% sensitivity and 100% specificity.

4. Finally, to complete the systematic analysis, application and presentation of curvelet texture features in CADx, we compared the classification performance of wrapping based discrete curvelet, Gabor filters and discrete wavelet texture features. This performance comparison is also necessary to measure the effectiveness and efficiency of curvelet texture features in CAD. Based on this

comparison, it can be decided whether this approach can be chosen as a standard for future CADx works.

6.2 Future Work

The future works appear from the limitations and the difficulties when we developed our system. The following developments can be made in the future:

1. We aim to build an interactive generalized CADx system for all medical image modalities.

2. To improve the accuracy results, the system has to take into account the feedback from the user.

3. To achieve further performance improvement of the system, we hope to optimize the system architecture and techniques that were used in this research. There exist some details setting can be discussed and optimized with the images classification issues.

4. The results demonstrate that each type of feature is effective for a particular type of images according to its contents, and using a combination of them gives better classification results. In future, we hope to try different types of features.

5. Using different benchmark image datasets with different semantics and categories to further evaluate our system and improve its limitations .

REFERENCES

[1] Yardimci, Ahmet. "Applications of soft computing to medical problems," In Intelligent Systems Design and Applications, 2009. ISDA'09. Ninth International Conference on, pp. 614-619. IEEE, 2009.

[2] L.M. Fletcher-Heath, L.O. Hall, D.B. Goldgof, F.R. Murtagh, "Automatic segmentation of non-enhancing brain tumors in magnetic resonance image," 43–63. Journal of Radiology,vol. 218, pp. 585-591,2001.

[3] S. Chaplot, L.M. Patnaik, N.R. Jaganbnathan,"Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network," Biomedical Signal Processing and Control, vol.1,issue 1, pp. 86-92, 2006.

[4] Lahmiri, Salim, and M. Boukadoum. "Brain MRI classification using an ensemble system and LH and HL wavelet sub-bands features," Computational Intelligence In Medical Imaging (CIMI), 2011 IEEE Third International Workshop On. IEEE, 2011.

[5] M. Tuceyran and A. K. Jain, "Texture analysis, in Handbook of Pattern Recognition and Computer Vision, World Scientific Publishing Co", Chapter 2.1,,pp.207-248, 1999.

[6] D. Gabor, "Theory of Communication," Journal of the Institution of Electrical Engineers, Vol. 93(26), 429-441, 1946.

[7] J. Y. Tou, Y. H. Tay, and P. Y. Lau, "Gabor Filters and Grey-level Cooccurrence Matrices in Texture Classification," MMU International Symposium on Information and Communications Technologies, 1-5 2007.

[8] D. Zhang, A. Wong, M. Indrawan, and G. Lu, "Content-based Image Retrieval Using Gabor Texture Features," IEEE Pacific-Rim on Multimedia (PCM00), 1139–1142, 2000.

[9] B. Kara and N. Watsuji, "Using wavelets for texture classification," in IJCI Proceedings of International Conference on Signal Processing, pp. 920–924, September 2003.

[10] S. Rajasekaran and G.A. Vijayalakshmi , "Neural Network, Fuzzy Logic and Genetic Algorithm," Prentice Hall of India, pg-13-20, 2003.

[11] W. Fourati, F. Kammoun, and M. S. Bouhlel, "Medical image denoising using wavelet thresholding," Journal of Testing and Evaluation, vol. 33, no. 5, pp. 364–369, 2005.

[12] Kharat, D. Kailash, P. Kulkarni, and M. B. Nagori. "Brain Tumor classification using Neural Network based methods," International Journal of Computer science and Informatics 1.4, 2012.

[13] Pemmaraju, Surya, et al. "Multiresolution wavelet decomposition and neuro-fuzzy clustering for segmentation of radiographic images," Computer-Based Medical Systems, Proceedings of the Eighth IEEE Symposium on. IEEE, 1995.

[14] Li, Lihua, et al. "Wavelet transform for directional feature extraction in medical imaging," Image Processing, Proceedings., International Conference on. Vol. 3. IEEE, 1997.

[15] Zhang, Yudong, et al. "Feature Extraction of Brain MRI by Stationary Wavelet Transform," Biomedical Engineering and Computer Science (ICBECS), International Conference on. IEEE, 2010.

[16] Othman, M. Fauzi, M. Ariffanan and M. Basri. "Probabilistic neural network for brain tumor classification," Intelligent Systems, Modeling and Simulation (ISMS), Second International Conference on. IEEE, 2011.

[17] Rathi, V. P., and S. Palani. "Brain tumor MRI image classification with feature selection and extraction using linear discriminant analysis," arXiv preprint arXiv:1208.2128, 2012.

[18] Wu, Cheng-Chi, et al. "A GA-based multiresolution feature selection for ultrasonic liver tissue characterization," Innovative Computing, Information and Control (ICICIC), Fourth International Conference on. IEEE, 2009.

[19] Rajini, N. Hema, and R. Bhavani. "Classification of MRI brain images using k-nearest neighbor and artificial neural network," Recent Trends in Information Technology (ICRTIT), International Conference on. IEEE, 2011.

[20] M. J. Fadili and J.-L. Starck, "Curvelets and Ridgelets," Encyclopedia of Complexity and System Science , in press., 2007.

[21] S. AlZubi, N. Islam, and M. Abbod, "Multiresolution Analysis Using Wavelet, Ridgelet, and Curvelet Transforms for Medical Image Segmentation," International Journal of Biomedical Imaging, vol. 2011, Article ID 136034, 18 pages, 2011.

[22] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying, "Fast Discrete Curvelet Transforms," Multiscale Modeling and Simulation, vol. 5, pp. 861-899, 2005.

[23] D. L. Donoho and M. R. Duncan, "Digital Curvelet Transform: Strategy, Implementation and Experiments," Stanford University, November, 1999.

[24] J.L. Starck, E. Candes, and D.L. Donoho, "The Curvelet Transform for Image Denoising," IEEE Transactions on Image Processing, 11(6), 670 -684, 2002.

[25] B. Belhaouari, Samir, F. Ibrahima and E. Meselhy, "A comparison of Wavelet and Curvelet for Breast Cancer Detection in Digital Mammogram," IEEE Transactions on Image Processing, 16(5), 555 -560, 2010.

[52] Kharrat, Ahmed, K. Gasmi, M. Messaoud, N. Benamrane and M. Abid, "Medical Image Classification Using an Optimal Feature Extraction Algorithm and a Supervised Classifier Technique." IJSSCI 3.2 (2011): 19-33. Web. 20 May. 2013.

[27] J.-L. Starck, M. Elad, and D. L. Donoho, "Image Decomposition Via The Combination of Sparse Rrepresentations and a Variational Approach," in IEEE Transactions on Image Processing, pp. 1570-1582, 2005.

[28] K. Kumar and Abhishek, "Artificial Neural Networks for Diagnosis of Kidney stone Disease," I.J. Information technology and computer science July-2012.

[29] M. N. Do, "Directional Multiresolution Image Representations," PhD thesis, EPFL,2001.

[30] F. E. Ali, I. M. El-Dokany, A. A. Saad, F. E. Abd El-Samie, "Curvelet fusion of MR and CT images," Progress in Electromagnitics Research, Vol. 3, pp. 215-224, 2008.

[31] M. Eltoukhy, I. Faye, B. Samir, "Using curvelet transform to detect breast cancer in digital mammogram," 5th International Colloquium on Signal Processing & Its Applications (CSPA), 2009.

[32] Johnson, J. Alex Becker, "The whole brain: Atlas," Harvard medical school, Second ed., 1999.

[33] Floares, A. Birlutiu, "Decision tree models for developing molecular classifiers for cancer diagnosis," Neural Networks (IJCNN), The 2012 International Joint Conference on , vol., no., pp.1,7, 10-15 June 2012.

[34] Shen, Liran, and Qingbo Yin. "Texture classification using curvelet transform," Proceedings of the International Symposium on Information Processing. 2009.

[35] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja. "Multiclass Brain Tumor Classification Using GA-SVM," In Proceedings of the 2011 Developments in E-systems Engineering (DESE '11). IEEE Computer Society, Washington, DC, USA, 182-187, 2011.

[36] Arizmendi, Carlos, et al. "Diagnosis of brain tumors from magnetic resonance spectroscopy using wavelets and Neural Networks," Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE. IEEE, 2010.

[37] K. Deepa , K.S. Gokul, K. Hinduja, R. Rajkumar, "An efficient approach to predict tumor in Brain image using classification techniques," Information Communication and Embedded Systems (ICICES), 2013 International Conference on , vol., no., pp.559,564, 21-22 Feb. 2013.

[38] Das, Arpita, and Mahua Bhattacharya, "A Study on prognosis of brain tumors using fuzzy logic and genetic algorithm based techniques," Bioinformatics, Systems Biology and Intelligent Computing, 2009. IJCBS'09. International Joint Conference on. IEEE, 2009.

[39] Mahamed, N.A., M.N. Ahmed and A. Farag, "Modified fuzzy c- mean in medical image Segmentation," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Piscataway, NJUSA, 6: 3429-3432, 2010.

[40] Hussain, S.J. Hussain, S.A. Raju, M.P. GiriPrasad, M. N. Satyanarayana, D. Venkatesh, "Neuro-fuzzy system for medical image processing," Emerging Trends in Robotics and Communication Technologies (INTERACT), 2010 International Conference on , vol., no., pp.52,55, 3-5 Dec. 2010.

[41] E.I papageorgiou, P.P Spyridonos, " Brain Tumor characterization using the soft computing technique of fuzzy maps"Applied soft computing 2008, vol 8,pp 820-828

[42] H. Rajini and Bhavani, "Automatic MR Brain Tumor Detection using Possibilistic   C-Means   and   K-Means   Clustering   with   Color

Segmentation," International Journal of Computer Applications 56(2):11-17, October 2012.

[43] Vrji, K. S. A., and J. Jayakumari. "Automatic detection of brain tumor based on magnetic resonance image using CADx System with watershed segmentation," Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), 2011 International Conference on. IEEE, 2011.

[44] Joshi, Dipali M., N. K. Rana, and V. M. Misra. "Classification of brain cancer using artificial neural network," Electronic Computer Technology (ICECT), 2010 International Conference on. IEEE, 2010.

[45] Bhardwaj, Anant, and Kapil Kumar Siddhu. "An Approach to Medical Image Classification Using Neuro Fuzzy Logic and ANFIS Classifier," International Journal of Computer Trends and Technology- volume4Issue3- 2013.

[46] http://www.curvelet.org/software.html.

[47] Zhang, Z. Dong, L. Wu, and S. Wang, "A hybrid method for MRI brain image classification," Expert Systems with Applications, Vol. 38, No. 8, 10049{10053, 2011.

[48] J.Jaya, K. Thanushkodi, "Segmentation of MR Brain Tumor Using Parallel ACO," International Journal of Computer and Network Security, Vol.2, No.6, pp. 150-153, 2010.

[49] S. Das, M. Chowdhury, and M. K. Kundu, "Brain mr image classification using multiscale geometric analysis of ripplet,"Progress In Electromagnetic Research, Vol. 137, 1-17, 2013.

[50] Zhang, S. Wang, and L. Wu, "A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO," Progress In Electromagnetics Research, Vol. 109, 325{343, 2010.

[51] Chaplot, S., L. M. Patnaik, and N. R. Jagannathan, "Classification magnetic resonance brain images using wavelets as input to support vector machine and neural network," Biomedical Signal Processing and Control, Vol. 1, No. 1, 86-92, 2006.