Brain Tumor Segmentation and Classification Using AI Techniques from MR Images

Muhammad Kaleem

154-FET/MSEE/F08

This dissertation is submitted as partial fulfillment of degree.

MS Electronics Engineering

DEPARTMENT OF ELETRONICS ENGINEERING

FACULTY OF ENGINEERING & TECHNOLOGY

INTERNATIONAL ISLAMIC UNIVERSITY

2010

M.phil
M-d

1. Diagnostic imaging – MR

Certification of Approval

It is certified that we have read the project report submitted by Muhammad Kaleem [154-FET/MSEE/F08]. It is our judgment that this project is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad for Master of Science Degree in Electrical Engineering.

External Examiner

Prof.Dr.Sayed Afaq Hussain,
Riphah International University,
I-14, Hajj Complex,
Islamabad.

Internal Examiner

Dr. Noaman Ahmed Khan,
Associate Professor & Chairman DEE,
FET, IIU, Islamabad.

Supervisor

Dr. M.Arfan Jaffar
Assistant Professor,
FAST-NU, Islamabad

Declaration

I certify that, the work done is of my own: the work has not been submitted previously, in whole or in the part, to qualify for any other academic award; the contents of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; and, any editorial work, paid or unpaid carried out by third party is acknowledged.

Muhammad Kaleem
154-FET/MSEE/F08

Acknowledgments

In the name of Allah, The Most Gracious, the Most Merciful.

All praise and glory goes to Almighty Allah, Who gave me the courage and patience to carry out this research work.

First and foremost, I would like to express my sincerest thanks to my parents for all their efforts and dedication since the time I came to this world, for me who can serve to his family and country. Thanks to my brother, wife, children, and sisters for their support and patience which always gave me the determination to finish this work.

I would like to express my gratitude to Dr.M.Arfan Jaffar, Dr.Ayyaz Hussain and Dr. Khurram Hameed for their guidance and encouragement during my Masters.

My special thanks go to Dr. Sajid Baloch and Akhtar Khan for their guidance and all the administrative support during my study period.

I extend my gratitude to all my close friends especially M.Ghufran, M.Shafiq, Wahid Bashir and Tahir Saleem for their encouragement and moral support.

Finally, thanks to my staff fellows Gulfraz, Amir and Saeed who helped me a lot during my thesis phase.

Muhammad Kaleem

154-FET/MSEE/F08

Abstract

Medical imaging makes use of the technology to noninvasively reveal the internal structure of the human body. By using medical imaging modalities patient's life can be improved through a precise and rapid treatment without any side effects. The main purpose of this dissertation is to develop an automatic method that can accurately classify a tumor from abnormal tissues. The images used for tumor segmentation and classification have been obtained from MRI modality. The objective of tumor detection and classification has been achieved by using segmentation and classification techniques. Discrete Cosine Transform and Discrete Wavelet Transform have been used for feature extraction. After feature extraction, feature reduction process has been achieved by using Principal Component Analysis. Brain MRI images have been classified by using Multi-Neural Network, Bayes Classifier, and Multi-Support Vector Machine. Segmentation of brain tumor has been done by using Morphological Watershed transform followed by graph based method. This computer diagnostic method has been tested on real data set for brain MRI images. This automatic method can help the physicians to detect and analyze pathologies leading for a more reliable diagnosis and treatment of brain related diseases.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

Brain Tumor detection and classification is one of the most challenging task in computer vision application. The complexity of this problem can be realized from the fact that animals and humans can recognize by constant learning from the beginning of their life. The animals and human can detect and classify things by making use of so many factors such as color, shape, edges, and from their previous knowledge. Humans can easily recognize between a chair and a table although both have similar properties. In this regard we can recognize by finding differences between the legs of table and chair, color, shape, and from our past experience. We can also differentiate among many other objects using other senses like, sense of touch and sense of smell, which are very difficult in image processing applications.

In case of tumor detection and classification there are assorted types of tumors. The diversity and tissue spread make tumor detection even difficult task for image processing engineers.

Most recently many tumor detection and classification systems have been developed for early detection of tumor before it starts to spread. Initially many biomedical imaging modalities have been improved to detect the tumor at early stages. Also many awareness methods among peoples to recognize early tumor symptoms have been introduced.

For automatic detection of tumor many computer aided diagnosis techniques have been developed. These techniques can help the physicians to non- invasively detect the tumor at an early stage. So in this dissertation we are going to build up a computerized diagnosis method for tumor isolation and classification. The main components of this method are segmentation, feature extraction and classification.

1.2 Segmentation of Brain Tumor

MRI segmentation has been used in this thesis for tumor diagnosis purposes. Many methods like edge based segmentation, region based segmentation, texture segmentation and Morphological Watershed segmentation methods have been utilized for brain tumor segmentation. The machine used to acquire the brain MRI images is given in the Figure 1.1.

The purpose of MRI segmentation has a host of numerous problems. For example, selection of most advantageous features is very crucial to maximize tissue dissimilarity discrimination. Moreover the level of machinist administration in the segmentation procedure will impact the strength of segmentation methods mainly in terms of inter class and intra class deviation. In this thesis we used watershed segmentation method for tumor detection. The problem with watershed is that of over segmentation which is resolved by using markers. The other important step in tumor detection is that of feature extraction and feature reduction. In the literature many other advanced methods of features extraction like fuzzy c mean clustering and k-means clustering have been used. There are also other approaches based on local features which are robust to variations in rotation, transformation and affine projections.

1.3 Feature Extraction for Brain MRI Images

Many existing methods have used local as well as global features with suitable kernel methods, but all have some sort of limitations, like additional steps to solve the problem of different dimension of feature vectors. Global features in combination with SVM and Fuzzy C mean have been used for tumor detection and classification. So the optimal selection of features is very important step in any computer aided system.

We in this thesis have used wavelet and discrete cosine transforms based features for tumor classification. There are large verities of tumor but our focus is to discriminate between benign and malignant tumor.

Feature reduction becomes an important step considering the dimensionality of given data is very large. One method of feature selection is based on texture of an image. In his method a small portion is selected and labeled as regions of significance. Huge quantities of textural features for tumor categorization has already been given in the literature and upon which discriminate examination were performed.

A new method based on SIFT [3] features can be used as future work for optimal feature selection and classification for brain tumor.

The flow of the main steps used in this thesis is shown in Figure 1.3.

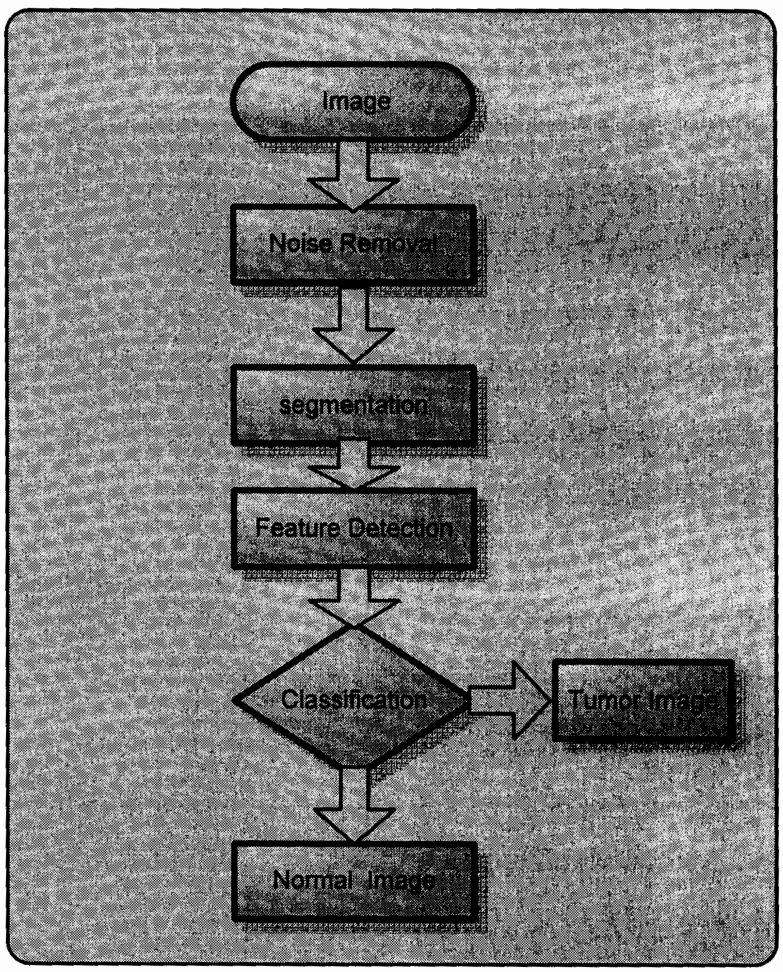


Figure 1.3 Main Architecture of Tumor Detection and Classification System

1.4 Classification of Brain Tumor

Image segmentation divides an image into different regions of similar attributes, but does not classify the tissue types. Classification is the process of labeling those distinct regions. There are many classification methods used some are uses supervised learning whereas some are based on unsupervised learning. In supervised learning the class information is known before hand. So both the data and associated is known before training is started.

Now after feature extraction and optimal feature selection the next task is the design of an efficient classifier. Both classifier design and optimal feature selection are very important in tumor detection process. The features extracted can be linear as well as non-linear. So the classifier must be efficient to solve both the problems. The differentiating characteristics of the classifiers are their ability to overcome the non-linearity inherent in the feature set.

Classification in general can be implemented in one of two ways.

- Quantitative performance analysis.
- Automated class assignments.

In quantitative performance method we analyze and then find differences between the features values obtained during PCA dimensionality reduction preserving maximum information.

In automated class assignment each distinct region is assigned a class and then classification accuracy is calculated from testing the data set.

We in this thesis have used Multi neural network and Multi support vector machine for classification purpose.

1.5 Organization of Thesis

Chapter 1: Chapter one gives over view of the whole dissertation giving brief introduction of each step.

Chapter 2: In chapter 2 we will give a through overview of literature survey of general segmentation methods. Dataset used for the assessment of the planned work of this thesis is also given in the same chapter.

Chapter 3: In chapter 3, we will in general explain image feature extraction techniques. Also brief overview of feature reduction and short summary of learning as well as non learning methods for classification of brain tumor will be discussed.

Chapter 4: In chapter 4, we will present a novel method based on Multi Networks based technique for the discrimination purpose of brain tumor.

Chapter 5: In chapter 5, we will discuss and suggest some work which can be done in future for feature extraction and tumor grading discrimination.

CHAPTER 2

MRI SEGMENTATION

2.1 Brief Introduction

Segmentation of an image entails the splitting up or separation of an image into regions of analogous distinctiveness. For monochrome images the most essential attribute for the segmentation purpose is the value of image luminance amplitude and color is the main component for color images. The edges within an image and texture are the other important attributes for the purpose of segmentation.

Other properties resembling gray level values and form may also help to separate an image into different regions of similar intra region properties and different inter regions characteristics.

The process of segmentation is only to divide an image into different regions it does not involve for the classification of individual segment. Furthermore, the segmentation process does not help to distinguish the entity segments nor it tells their relationship to any other part of the image.

Broadly speaking there is no theory behind the image segmentation. As a result there is no particular technique of image segmentation. There are numbers of ad hoc process that have been used largely and have received popularity to some extent.

As the image segmentation methods are ad hoc, there should be some means of assessing the performance of these methods. Following are the qualitative guide lines for high-quality image segmentation.

"Regions of image for the process of segmentation should be consistent and homogenous with value to some characteristics such as gray tone or texture. Regions interiors should be uncomplicated and with no many small holes. Neighboring regions of segmentation should have considerably dissimilar values with respect to the feature on which they are identical. Margins of every segment should be simple, not tattered, and have to be spatially precise" [4]. Unfortunately, so far in the given literature there is no quantitative performance metric for image segmentation yet has been developed.

2.2 General Segmentation

Image segmentation techniques are normally based on host of properties but fundamentally two essential characteristics of intensities which are discontinuity and similarity are of utmost important. For a lot of image analysis and for application point of view segmentation of the given image is the basic important procedure. The ultimate result of the image processing system is governed by the segmentation techniques; if the segmentation is accurately done then the rest of the task becomes easier. So in the literature there are number of methods available but in this dissertation we will talk about the most fundamental ones. Some of the general methods of segmenting an image are discussed in the subsequent sections.

- Thresholding base image segmentation.
- Region segmentation procedures.
- Neural Network based segmentation.
- Segmentation methods based on edges of images.
- Seed growing segmentation methods

2.2.1 Segmentation Methods Based on Thresholding

Image Thresholding is the most commonly used technique for segmenting an image. This is because that Thresholding is easy to implement and have some intuitive properties.

There are many images that can be discriminate as containing a number of objects of attention of rationally consistent brightness on dark surroundings. The most common examples include microscopic biomedical samples, typewritten and handwritten text, and airplanes on the runway. In such a situation the object and surroundings pixels have gray levels grouping into two leading modes. So the most immediate choice for extracting these objects from background is to choose an appropriate threshold T that divides these objects from the back ground. This is one of the simple cases of image Thresholding; there are additional general cases of this approach where many central modes characterize an image. These dominant modes methods are called as multilevel Thresholding.

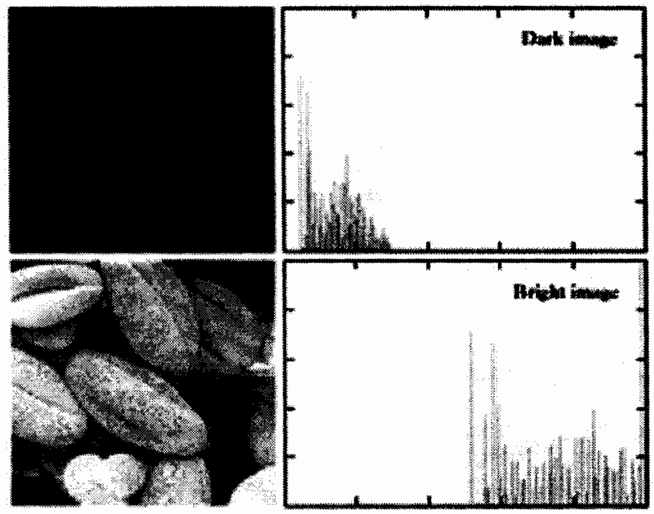


Figure 2.1 Single level Thresholding [5]

Mathematically the process of Thresholding can be defined as given in the equation below.

$$g(x,y) = \begin{cases} 1 \text{ if } f(x,y) > T \\ 0 \text{ if } f(x,y) \leq T \end{cases} \qquad 2.1$$

In equation 2.1 pixels labeled 1 corresponds to objects whereas pixels labeled 0 corresponds to the background. The value of T depends on the problem under consideration and this process is generally called as global Thresholding.

Thresholding can also be viewed as an operator which as the form given below which also describe the local properties of images as well.

$$T = T(x, y, p(x, y), f(x, y)) \qquad 2.2$$

In equation 2.2 f(x , y) is the gray level of point (x , y) and p(x , y) denotes some local properties of that point e.g. average gray level of a neighborhood centered at (x , y). If the value of threshold depends both on the f(x, y) and the local properties the resulting Thresholding method is called as local Thresholding.

The value of T can also be determined on the spatial coordinates x and y, the threshold is called as dynamic or adaptive.

The process of multilevel Thresholding can be explained in the form of equation given below.

$$g(x,y)=\begin{cases} obj\ 1 & \text{if } T1 < f(x,y) \leq T2 \\ \text{obj2} & \text{if } f(x,y) > T2 \\ \text{Background} & \text{if } f(x,y) \leq T1 \end{cases} \quad 2.3$$

The values of T1 and T2 can be completely determined on the basis of problem under considerations.

In the literature there are many methods of Thresholding some of the most common methods are given below.

- Thresholding based on EM algorithm [6].
- Thresholding based on clustering methods [7].
- Thresholding based on the Fuzzy entropy.
- Thresholding based on attribute similarity method.
- Binary Swarm optimization Thresholding [8].
- Co-occurrence based Thresholding method [9].

2.2.2 Region Based Segmentation Methods

The main objective of segmentation is to partition the image into different regions based on some predefined criterion functions. In this method segmentation is done by directly finding the regions.

There are some fundamental properties of segmentation which must be satisfied for accurate segmentation which are given in the equation form below.

$$\bigcup_{i=1}^{n} Ri = R \qquad 2.4$$

$$R_i \text{ is a connected region,i=1,2....n} \qquad 2.5$$

$$R_i \cap R_j = \varnothing \text{ for all i and j, } i \neq j \qquad 2.6$$

$$P(R_i) = TRUE \text{ for i=1,2...n}$$
$$P(R_i \cup R_j) = \text{FALSE for } i \neq j \qquad 2.7$$

For the accurate segmentation all of the above properties must be satisfied. In the above equations $P(R_i)$ is the logical predicate for all the points in the set R_i. The pictorial view of image segmentation based on color is given in the Figure 2.2 and Figure 2.3.

Figure 2.2 Original image for segmentation

Figure 2.3 Resultant image based on color segmentation

Region based segmentation can be divided into main three types.

- Region growing methods.
- Region splitting and Merging.
- Region segmentation based on fuzzy sets features [10].

2.2.3 ANN Based Segmentation

Neural Networks are the supervised learning technique used for the segmentation. Artificial Neural Network (ANN) is a very common and useful supervised machine learning technique. Basically Neural Networks model the brain of human for solving many complex problems in the real world.

For the purpose of segmentation the features of the image are extracted by using many of the familiar techniques. Once the features are extracted their performance is evaluated based on some metric and then these features are passed through the feature reduction process. In feature reduction process the maximum information is preserved in the low dimensional space while discarding the redundant features. Once optimal features have been extracted these are given at the input layer of NN. The processes of Neural Network automatically partition the image into different segments.

ANN is playing an integral part in the image classification and segmentation [11]. A large number of ANN has been well-known for image segmentation like, Feed forward back propagation artificial neural network, Self Organizing neural network, Hopfield based networks; Radial basis functions based network, and Kohonen network architectures.

The essence of neural networks is the use of multitude of elemental nonlinear computing elements called as neurons. These neurons are interconnected in the similar fashion as the neurons in the brain. The process of neural networks can be portioned into two categories i.e. learning and testing. In the learning phase the network is trained as human do and next in the testing phase novel patterns are given to the network for recognizing.

The most basic formulation of the NN can be explained in equation form which is given below.

$$d(X) = \sum_{i=1}^{n} w_i x_i + b \qquad 2.8$$

In equation 2.8 W is called as weight calculation matrix that is calculated during the training phase. The term b is called the bias and x is the input to the input layer of the neurons. Equation 2.8 can be expanded for illustration purpose as:

$$w_1 x_1 + w_2 x_2 \ldots\ldots\ldots\ldots\ldots w_n x_n = 0 \qquad 2.9$$

If we use only the global Thresholding method for segmenting the any given input image than we can represent equation 2.8 as:

$$o = \begin{cases} +1 \text{ if} \sum wixi > -b \\ -1 \text{ if} \sum wixi < -b \end{cases} \qquad 2.10$$

The output of each layer is accomplished by using an activation function which depends upon the application. There are basically three stage of any neural network one is called the input layer, the hidden layer and the output layer. There can be more than one hidden layer in any neural network. There can be different number of neurons per layer. The most widely used form of equation 2.8 in terms of matrix algebra can be formulated as:

$$d(Y) = W^T Y \qquad 2.11$$

2.2.4 Segmentation Based on Edges

Edges are the most important point in an image to segment. False edge can be detected due to noise and inconsistency in the Thresholding values. The process of segmenting an image based on edge detection also pose problems due to above and below segmentation, also due to the presence of unwanted noise and differing methods in the Thresholding methods [7]. Edge can be detected by looking at discontinuities in the image. Edge finding can be shared with morphological filtering methods for segmenting any given image.

Many different methods based on first derivative and second derivative can be used to detect different edges in horizontal and vertical directions. There are three kinds of discontinuities of intensity. Edge detection can be used for ramp, step and sudden jump like changes in an image. Different masks of variable size can be used for detecting the edges. False edges can be detected due to noise in the input image. Therefore it is mandatory to apply noise estimation techniques for the purpose of noise removal. Following is the list of some of the most widely used edge detection methods.

2.2.4.1 Robert Edge Finding Scheme

Edges in this method can be finding using different kernels in the horizontal as well as vertical directions. In this method kernels are useful to the novel images to be edge detected and we get the consequential edge founded image.

$$Rx = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } Ry = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad 2.12$$

In above written kernel Rx is used for the finding of horizontal edges in the images and the kernel used in Ry can be applied for finding the vertical edges in the given images. If we characterize the given image by *I* then we are able to get the resultant images as mentioned below.

$$Ex = Rx * I \qquad \text{and} \qquad Ey = Ry * I \qquad 2.13$$

The average edge image E can be calculated as under

$$|E| = \sqrt{Ex^2 + Ey^2} \qquad \text{and} \qquad \theta = \tan^{-1}\left(\frac{Ey}{Ex}\right) \qquad 2.14$$

2.2.4.2 Edge Detection Using Sobel Operator

Sobel operator is one of the useful methods for detecting edges in the images. So following are the two operators which can be applied in vertical as well as horizontal edge detection.

$$Sx = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \text{ and } Sy = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \qquad 2.15$$

Where in the above equation Sx used for horizontal edge detection and Sy is used for vertical edge detection.

$$Ex = Sx * I \quad \text{and} \quad Ey = Sy * I \qquad 2.16$$

The standard edge image *E* can be considered as given below

$$|E| = \sqrt{Ex^2 + Ey^2} \quad \text{and} \quad \theta = \tan^{-1}\left(\frac{Ey}{Ex}\right) \qquad 2.17$$

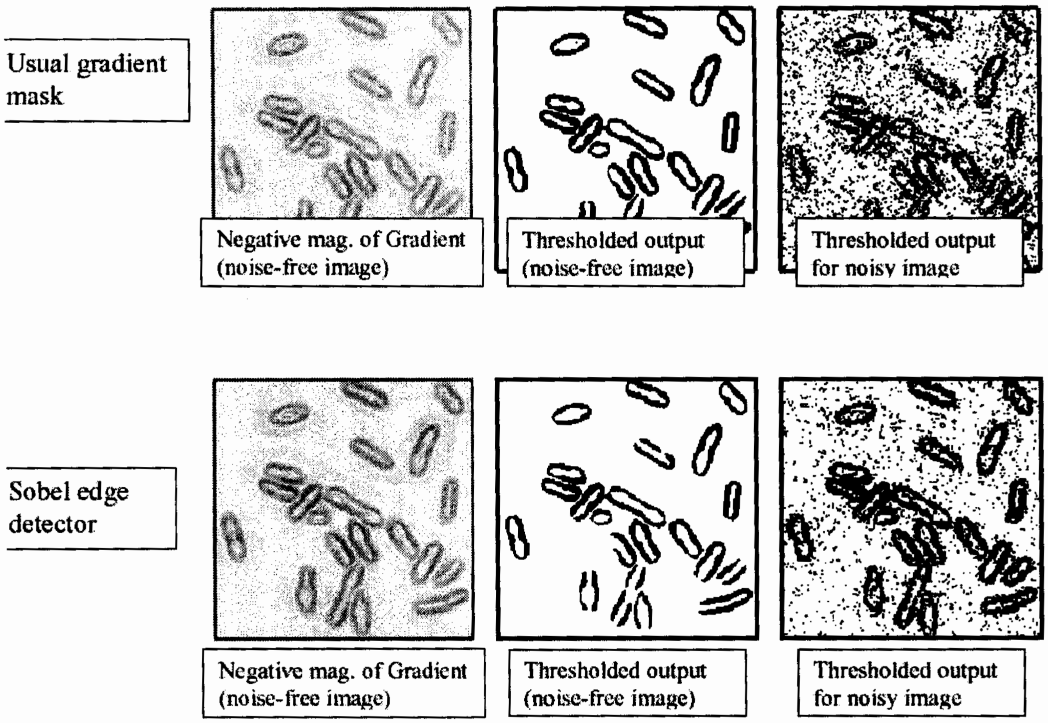


Figure 2.4 Examples of Sobel edge detector

2.2.4.3 Edge Detection Using Prewitt Kernel

For the purpose of edge detection in this method following filters are used both for horizontal as well as horizontal edges in any given image.

$$P_x = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad Py = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix} \qquad 2.18$$

$$Ex = Px * I \qquad \text{and} \qquad Ey = Py * I \qquad 2.19$$

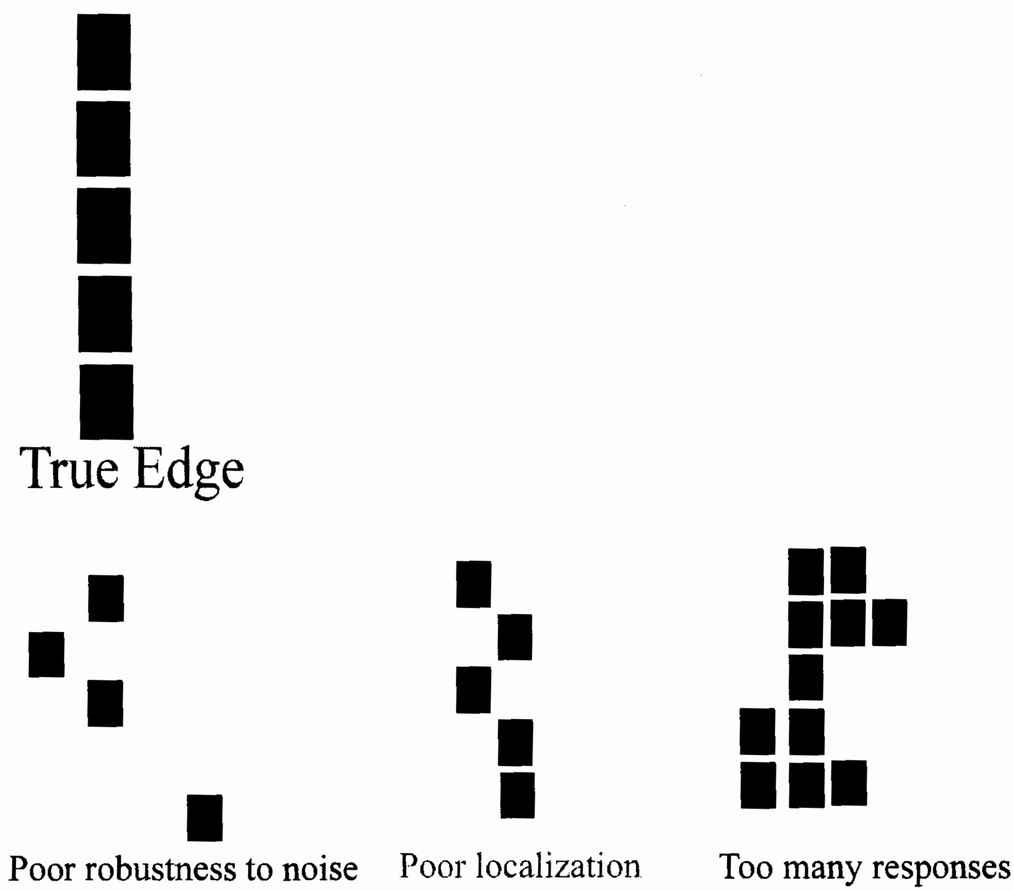
The average edge image E can be computed as given below

$$|E| = \sqrt{Ex^2 + Ey^2} \quad \text{and} \quad \theta = \tan^{-1}\left(\frac{Ey}{Ex}\right) \qquad 2.20$$

2.2.4.4 Edge Detection Based on Canny Filters

There are some fundamental requirements for good edge detector and these are listed below.

- Detector must be robustness to noise and must response to true edges.
- Localization is the other important factor.
- Too many or Too less responses.



A god detector must be in compliance to these guide lines. A detector must give thin edge as compare to thick edges.

There are some other criterions for good edge detector.

- Signal to Noise Ratio: The optimal detector should respond more strongly to true edges as compared to noisy edges.
- Good Localization: The edges detected have to be as close to true edge as likely.
- Single event response limit: The detectors have to return one point only for every edge point.

Canny modeled the above two criterion and the constraint by mathematical parameters. He tried to maximize the SNR and localization parameters subject to the constraint defined mathematically. An efficient implementation of the filter which gives required optimality turns out to first order derivative of the Gaussian functions. Canny edge detector includes the following.

- Convolution with derivative of Gaussian.
- Non-maximum Suppression.
- Hysteresis Thresholding.

First of all the novel image is get smoothed by using the Gaussian as:

$$S = G_\sigma * I \tag{2.21}$$

Where G_σ is given by

$$G_\sigma = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2.22}$$

Secondly compute the x and y derivatives as:

$$\nabla S = \left[\frac{\partial}{\partial x} S \quad \frac{\partial}{\partial y} S \right]^T = \left[S_x \quad S_y \right]^T \tag{2.23}$$

Thirdly compute the gradient magnitude and orientation as:

$$|\nabla S| = \sqrt{S_x^2 + S_y^2} \quad 2.24$$

$$\theta = \tan^{-1}\frac{S_y}{S_x} \quad 2.25$$

$$\nabla S = \nabla(G_\sigma * I) = \nabla G_\sigma * I \quad 2.26$$

$$\nabla G_\sigma = \left[\frac{\partial G_\sigma}{\partial x} \quad \frac{\partial G_\sigma}{\partial y}\right]^T \quad 2.27$$

$$\nabla S = \left[\frac{\partial G_\sigma}{\partial x} * I \quad \frac{\partial G_\sigma}{\partial y} * I\right]^T \quad 2.28$$

The outputs of the non-maximum suppressed image are edges but since they are threshold with single value there may be disconnected edges. To overcome this problem researchers have used Hysteresis Thresholding. Following are the main steps in this method.

- If the gradient at a pixel is above High declare it an edge pixel.
- If the gradient at a pixel is below Low declare it a non-edge pixel.
- If the gradient at a pixel is between Low and High then declare it an edge pixel if and only if it is connected to an edge pixel directly or via pixels between Low and High.

2.3 Image Acquisition Modalities

So far different techniques in medical world have been used for revealing the internal structure of the body organs with negligible invasiveness. However, patient protection receive main concern when scheming biomedical imaging strategy, with the major challenge being able to get the clearest and a large amount explanatory image of an organ with some degree of side effects.

There are a number of different medical imaging modalities which can be used to get insight into the human body. Some of those are listed below.

- Ultrasound imaging method
- Computed tomography
- Positrons emission tomography
- Single photon emission computed tomography
- Magnetic resonance imaging
- Functional MRI (fMRI)

2.3.1 Magnetic Resonance Imaging – MRI

MRI is a noninvasive technique used in medical imaging for diagnosing and prognosis purposes. In MRI technique a non-ionizing is applied for the purpose of getting the cross sectional images of soft tissues in the human body. Almost 75% of the human body is composed of water as hydrogen is in excessive amount in the body. In MRI a magnetic field is applied on the human body to align proton in parallel fashion. These protons are then exposed by strong but harmless radio waves that will force these protons to initial random position. As soon as these proton align themselves back they will produce a radio signal which is detected by a receiver in the MRI scanner. After this procedure images can be captured in the computer for further processing and analysis purposes. The scanner machine and normal brain tumor are given below.

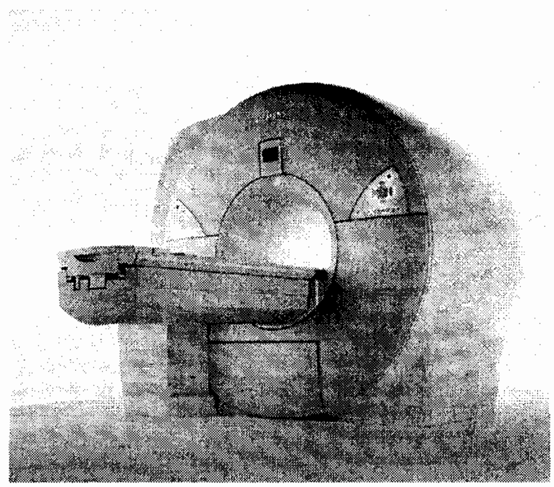


Figure 2.5 MRI Scanner Machine [12]

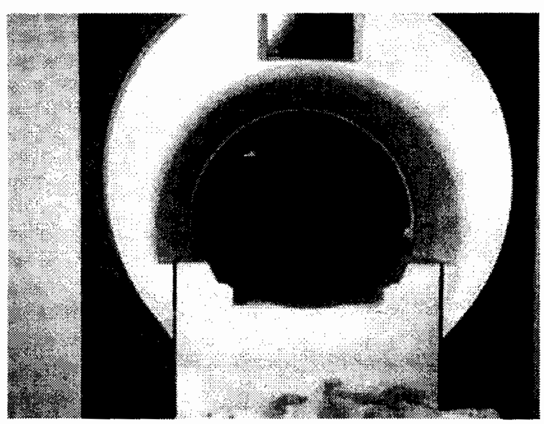


Figure 2.6 Inside view of MRI Machine [13]

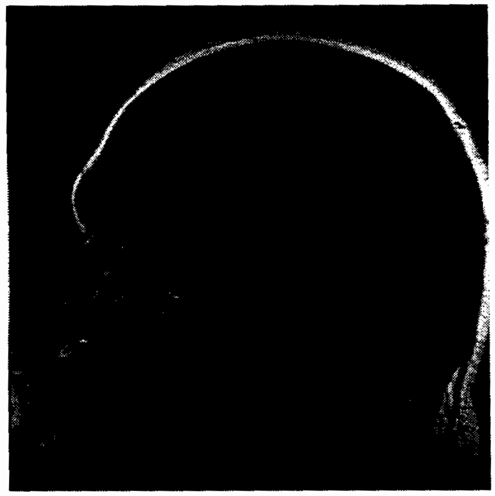
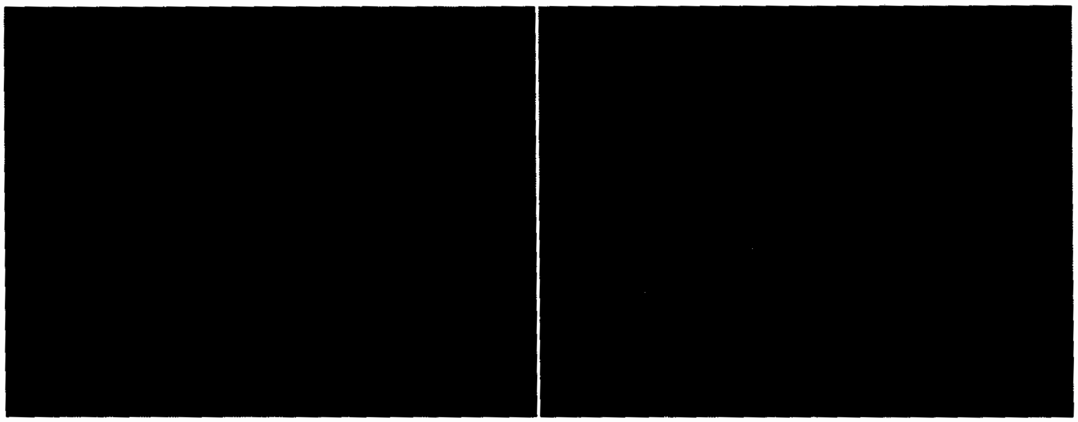


Figure 2.7 Normal brain image

Figure 2.8 Brain image with tumor

2.4 Overview Of Brain Tumor

Survay of WHO shows that malignat tumor are the leading cause of death world wiede. There have been many efforts towards the understanding of this deadly disease. There are so many awarness methods and programs have been launched to cure this disease. There are many imaging techniques which have been developed for early detection of tumor.



a)Normal Brain Cells b) Tumorous Brain Cells

Figure 2.9 Figure of Normal and Tumorous Brain Cells

CHAPTER 3

BRAIN FEATURE DETECTION AND CLASSIFICATION METHODS

3.1 Overview and Introduction

In our daily life we try to classify things base on the distinctive properties inherent to the things. As an example we can classify between a chair and a table based on the contour of the chair and table. So for this case shape is the important feature to distinguish between the two. Similarly there are other shuch important features which are natural as well as artifical used to classify things from one another.

In the image prosessing domain An image feature is a distinctive basic characteristic or quality of an image. Various features can be natural while other artifical , because artifical features result from specific manipulation to the image under processing.

Feature detection is the most fundamental step in almost every computor vision application. So proper detection of all the features is very important since the performancce of the whole system is largely depends on these primitive features. Therefor to enhance the reliability of computor vision application features must be unique and ccarrying maximum information in the given images. Further features can be natural ,artifical , local and global. The luminance values in the image and the values of texture gray scale are the natural features. The artifical features include image amplitue histogram and spatial frequency spectra. Furthermore features are of amjor imortance for image segmentation and image classification. But the conceptual boundary between feature extraction and classificartion is some how very much difficult. An ideal feature extracture would make the job of classifier trivial conversaly a powerful calssifier would make the job of feature extracture less important. So this characteristic is upon the user fro realistic rather than therotical reasons.

Feature can be:

- Edges
- Corner points
- Texture

Image classification can be separated in to two basic steps. Accurate feature extraction and classification. Both of these steps are of fundamental imortance in any vision application. Features can be local and globl features. Initially people used global features for image recognition and classification but recent research shows that local features are most important for object recognition and classification.

So far in the text many feature detection techniques have been developed e.g. texture features, gabor features , feature based on Discrete wavelet transform methods , minimum noise transform, discriminant mthods, PCA, decision boundary based feature extraction, non-parametric feature detection and spectral mixture analysis methods have been used in many different applicatiions.

3.2 Feature Detection and Extraction

Feature detection is a transforming function y = f(x): $R^d = R^{d'}$ with $d' < d$, such that the new and compact dimensional vector y keep as much of the given characteristics in the data x as given in the transformation below.

$$\boxed{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \rightarrow \begin{bmatrix} y_1 \\ \vdots \\ y_{d'} \end{bmatrix} = f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}\right)}$$

Extraction if features from an image is the most important and basic task in any image processing appliction. Good and optimal features play fundamental role in object recognition and classification applications. So feature extraction canbe through linear as well as non-linear methods. In [14] Yong et.al has suggested feature ectraction and detection which is based on linear separable methods. This method makes use of Fisher method. The data which is used by this method can be normal distributed as well multi modeled distribution.

Other feature extraction techniques seggested by by Gil et al [15] are based on stochastic learning methods. Gede et.al in [16] has suggasted a method for facial recognition which is based on multiresolution and frequency features metric. Another method proposed by Asako et .al [17] gets features based on similarity measure between two images.

Some other researcher like Haralick et al [18] has also proposed another method for optimal feature detection which is used in numerous object recognition and classification applications.

3.2.1 Local Features Vs Global Features

The process of feature extraction can be dividing in to two main classes' i.e. local features and global features [3]. Both have been used by researches for the segmentation and classification purposes. Global feature extractor takes the whole image as input and then process the image for classification purposes. Local features are very important in the literature because of the similarity to the human visual system. This similarity of local feature extraction to human visual system allows for accurate recognition and classification of objects under different conditions. Local features are very robust to the illumination variation, scale variation, rotation and affine transformations. These important properties of local features make them useful for application to many computer vision applications.

A local feature in an image can be detected by an interest point detector. This interest point can be corner point, an edge point or a corner point detected by applying the SIFT algorithm. There are other algorithms like Harris corner point detector which can also be used for extracting local features. These two algorithms are widely used along with Gabor algorithm for local feature extraction.

An overview of some of the basic techniques used for feature detection and extraction methods used widely in image processing application is listed in the literature given below.

3.3 Discrete Cosine Transform

DCT is the one most basic techniqe used for the purpose of feature extraction in computer vision applications. Although it gives good result for image compression it can aslo used for extracting features. When implementing in MATLAB we use dct2 function to take out features from given MRI images. It is a technique which is used to take out global features from an image. DCT can be used for transforming the signal into frequency domain. Hazim et.al [19] has suggested feature extraction by using DCT for face recognition purpose. DCT can transform the signal into fewest coefficients without degradation of the origional signal. Symmetry and separability are the two most important properties of DCT. The most basic and important equations for implementing DCT are listed in[20][21].

$$C(u,v) = \alpha(u)\alpha(v)\sum\sum f(x,y)\cos\left[\frac{u\pi(2x+1)}{2N}\right]\cos\left[\frac{v\pi(2y+1)}{2N}\right]$$
$$0 \le u \le N \;\&\; 0 \le v \le N \tag{3.1}$$

$$\alpha(u) = \begin{cases} \sqrt{1/N} \text{ for } u=0 \\ \sqrt{2/N} \text{ for } u = 1,2,3.....N-1 \end{cases}$$

We have used MATLAB builtin function to calculate the DCT of the given MRI images. To avoid computational complexity we have most relevant frequency features for the classification of MRI images.

3.4 Discrete Wavelet Transform

Multiresolution or multiscale analysis is a fine technique we there is a need to analyze the signal at different level of resolution. So a signal is decomposed into sub level of resolution and examined at different frequency levels[22]. In pattern recognition high contrast signals are analyzed at low level of resolution[23][24], while high level resolution is more appropriate for low contrast or small size objects. Multiresolution analysis can be divided in to twomain methods and are called as wavelet decomposition and steerable pyramids.

Discrete wavelet transform has been obtained from continuous wavelet transform. So continuous vavelet transform makes use of some fundamental fuctions called as mother wavelet. Continuous wavelet transform can be obtained by the combinition of following functions like ψ(x) and $f(x)$, which are defined in following equations.

$$W_{\psi}(p,q) = \int_{-\infty}^{\infty} f(x)\psi_{p,q}(x)dx \qquad 3.2$$

Where

$$\psi_{p,q}(x) = \frac{1}{\sqrt{p}}\psi\left(\frac{x-q}{p}\right) \qquad 3.3$$

Where in the above equations p is called the scaled parameter and q is called as translation parameters, respectively.

Furthermore wavelet decomposition provides analysis both in the time domain and in the frequency. Each level of breakdown contains the information of a particular level and direction. 2-D Wavelet decomposition can be obtained by using different filter banks as:

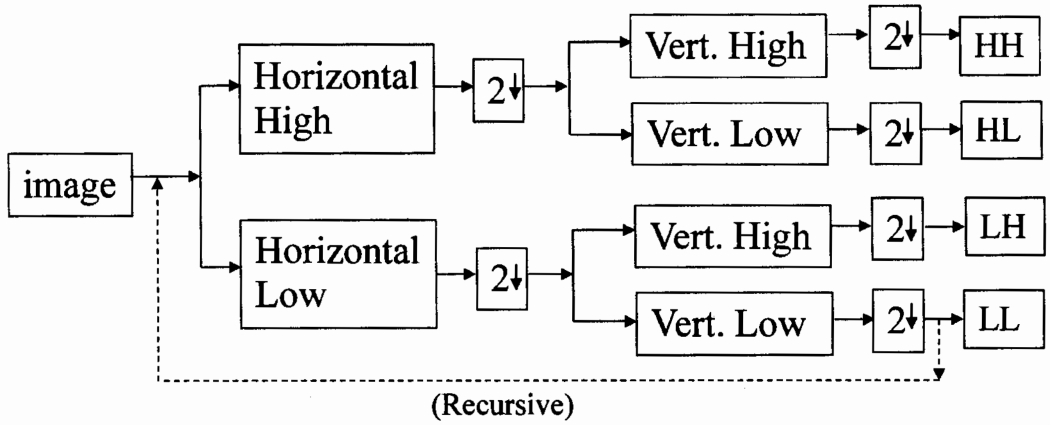


Figure 3.1 Two Dimensional Wavelet Analyses

The above decomposition can be shown pictorially as:

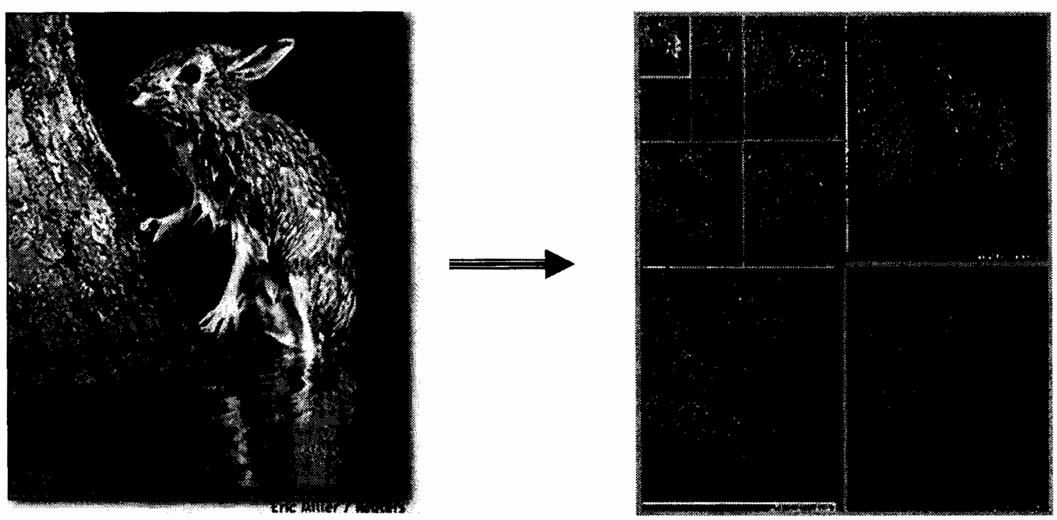


Figure 3.2: Input image and its Wavelet Decomposition

For the analysis of an image in two directions, we need a scaling functions $\varphi(x,y)$ in two dimensions and we are also required three two dimensional wavelets which are given as $\psi^H(x,y), \psi^V(x,y), \psi^D(x,y)$.

The separable function is

$$\varphi(x,y) = \varphi(x)\varphi(y) \tag{3.4}$$

And separable, "directionally sensitive" wavelets are given in the following equations.

$$\psi^{H}(x,y)=\psi(x)\varphi(y) \quad 3.5$$

$$\psi^{V}(x,y)=\varphi(x)\psi(y) \quad 3.6$$

$$\psi^{D}(x,y)=\psi(x)\psi(y) \quad 3.7$$

Then the DWT of function $f(x,y)$ of size M X N can be represented in an equation form as given below.

$$W_{\varphi}(j_o,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\varphi_{jo},m,n(x,y) \quad 3.8$$

$$W_{\psi}^{i}(j,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\psi_{j,m,n}^{i}(x,y) \quad \mathbf{i=\{H, V, D\}}\ 3.9$$

Where in the above equations $W_{\varphi}(j_0,m,n)$ [5] coefficients define an approximation of $f(x,y)$ at scale j_o and $W_{\psi}^{i}(j,m,n)$ [5] coefficients are used to add vertical, horizontal, and diagonal details for scales $j \geq j_o$. So a two directional DWT can be implemented by digital filters and down samplers and with two directional scaling and wavelet functions. The approximation, vertical, horizontal and detailed components are shown in the figure below.

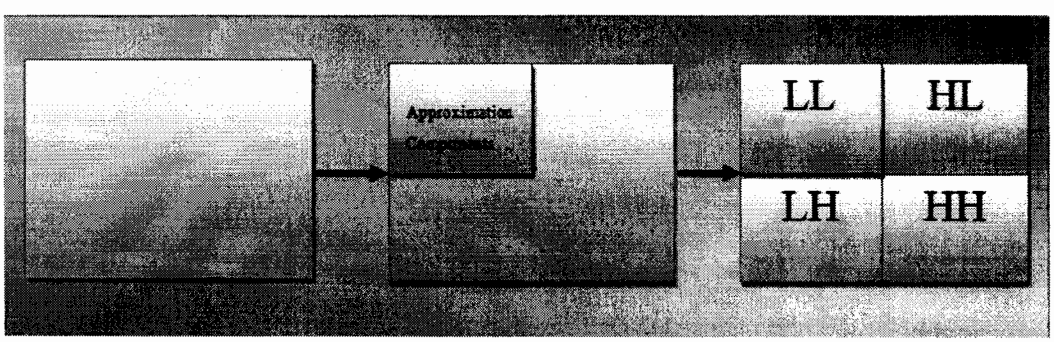


Figure 3.3 Image and Its DWT components

Similarly there is an inverse relationship between time and frequency domain. This relationship can be shown in the following figure. In the figure the horizontal direction shows time and the vertical axis represents the frequency axes respectively. In this we can see that wavelet transform can be used to analyze the data at various level of resolution and thus gives detail insight of the signal under considerations.

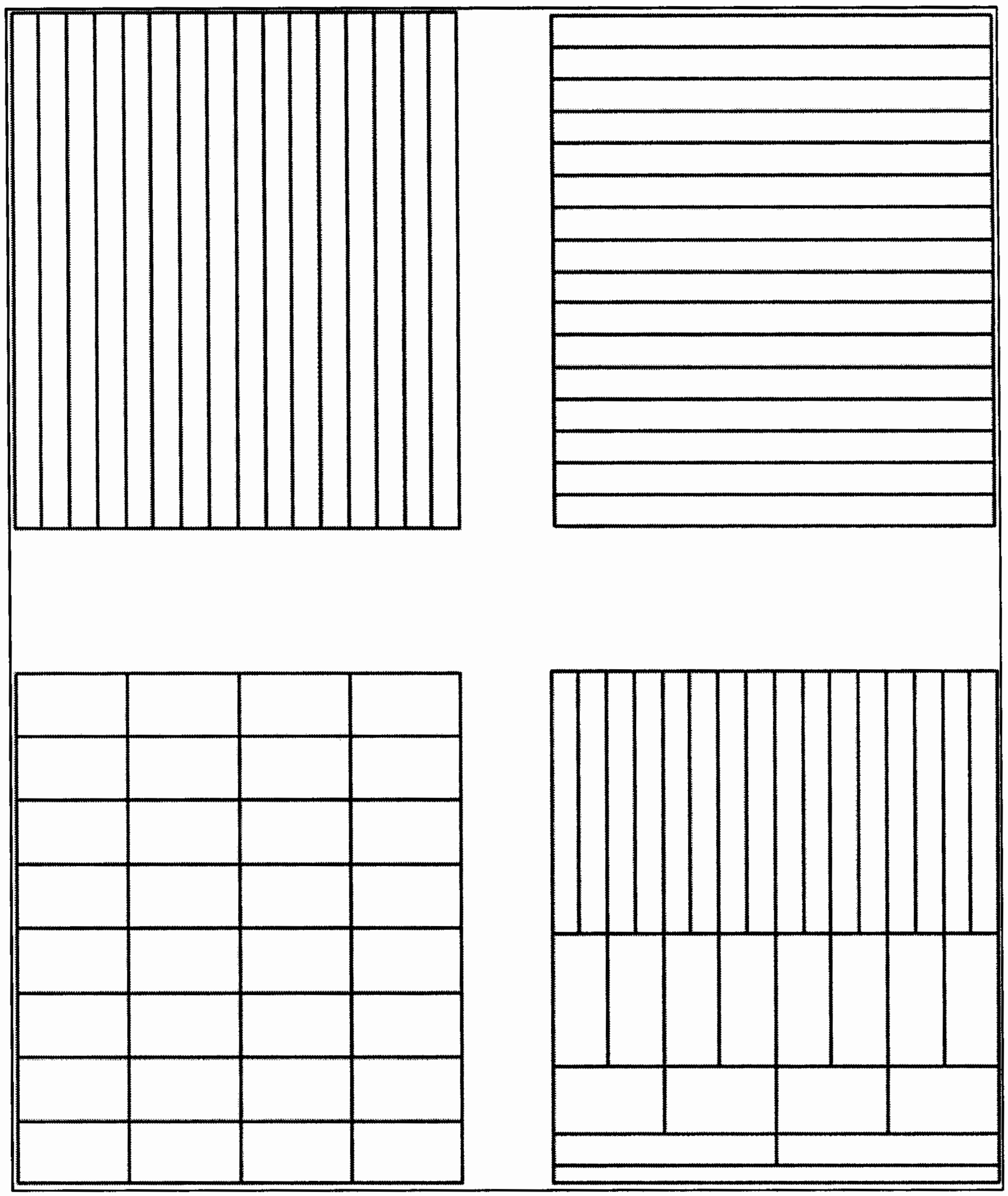


Figure 3.4 Time Frequency Tiling for (a) Top left: for sampled data (b) Top Right: Fourier Transform (c) Bottom Left: Windowed Fourier Transform and (d) Bottom Right: Wavelet Transform [25].

3.5 Texture

An important approach for image analysis in computer vision application is the use of textures. Even though there is no mathematicl defition of texture, naturally this method makes use of certain properties like smoothness, roughness, and regularity. Furthermore another defination of texture is "Repitition of patterns or patterns over a region of sufficiently large size." Textures are made up of distinct and small subelements also callled as textons orimage primitives repititing overa a region in a meaningfull patterns. Texture is the property of sufficiently large regions. The examples of texture is shown in the figure below.

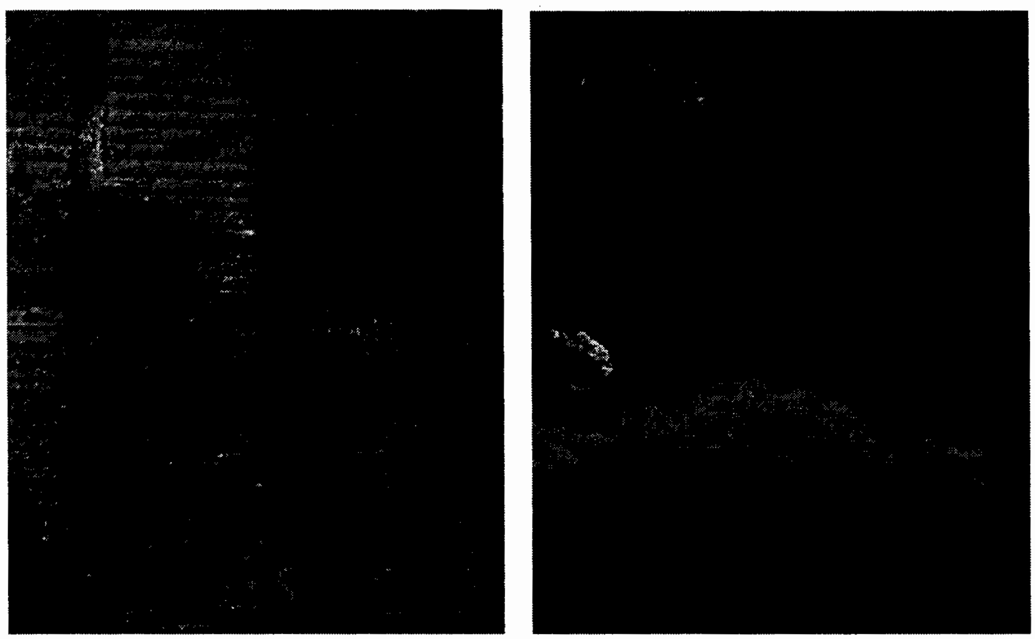


Figure 3.5 Texture Example of Natural Scenes.

Textures can also be used for synthesis purposes i.e. making whole from the parts. If given a finite sample of textures the goal is to synthesize the other samples from those same samples. The most fundamental property of texture is that the sample must be of sufficiently large size and example is given below.

There are three basic methodologies used in image processing to illustrate the texture of a region. The basic sub elements of texture are bars and spots. Texture is an important region description and gives efficient result depending on the application. In computer vision application texture methods can be divided into following three different categories.

- Statistical Methods
- Spectral Methods
- Structural Methods

Statistical approaches are used for the classification of textures such as smoothness, coarse, grainy, and so on.

Structural methods are suitable with the arrangements of image primitives, such as the explanation of texture based on a regular basis spaced similar lines.

Spectral methods are based on the characteristics of the Fourier transform and are used mainly to detect overall periodicity in an image by identifying high energy, fine peaks in the range.

3.5.1 Features Based on First-order histogram

The most simplest method used for describing texture is the statistical moments of the gray level histogram of an image or region.

Following are the most significant features one can get by using technique based on image histogram[26].

mean $$\mu = \sum_{i=1}^{G-1} iP(i) \tag{3.10}$$

Variance $$\sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 P(i) \tag{3.11}$$

Skewness: $$\mu_3 = \sigma^{-3} \sum_{l=0}^{G} (l-\mu)^3 p(i) \tag{3.12}$$

Kurtosis: $$\mu_4 = \sigma^{-4} \sum_{l=0}^{G} (l-\mu)^4 p(i)^{-3} \tag{3.13}$$

Energy: $$E = \sum_{l=0}^{G} [p(i)]^2 \tag{3.14}$$

Entropy: $$H = -\left(\sum_{l=0}^{G} p(i) \log_2[p(i)] \right) \tag{3.15}$$

3.5.2 Features from Co-occurrence Matrix

Somer textural based features can be obtained based on the combined amplitude histogram pairs of pixels. In any given iamge if the region has fine texture, then the two-directional histogram of pixels will give uniform result, and for the coarse texture the histogram values will be at an angle towards the diagonal direction of the histogram.

For the calculation of co-occurance matrix we choose a SxS window where each pixels is quantized over a certain range. We can count the gray level valuses in the specified window with certain angle and at a certain distance.

Let $S(i,j,d,\theta)$ represent an image where i,j are the gray level values at distance d=1 and at $\theta = 0^0$ the given image and the resultant co-occurance matrix is shown below:

50	51	52	50
53	51	51	52
51	50	51	52
52	53	53	52

	50	51	52	53
50	0	2	0	0
51	1	1	3	0
52	1	0	0	1
53	0	1	1	1

Table 3.1 Intensity image above and its co-occurrence matrix with d=1 and theta=0

The simple Tabular calculation of an input image of 5x5 in four direction and distance d=1 is shown below.

Original image

1	1	1	0	1
1	3	3	0	2
1	2	3	2	2
2	2	2	3	0
2	4	2	4	1

$\theta = 0^o$

0	1	1	0	0
1	2	1	1	0
0	0	3	2	2
2	0	1	1	0
0	1	1	0	0

$\theta = 45^o$

0	1	1	0	0
0	0	0	1	0
2	2	2	1	0
0	2	0	2	0
0	0	2	0	0

$\theta = 90^o$

1	0	1	0	0
1	2	0	0	0
1	2	4	2	0
0	2	1	1	0
0	0	1	1	0

$\theta = 135^o$

0	1	0	0	0
0	1	0	1	0
0	0	4	3	0
2	1	1	0	0
1	0	1	0	0

Table 3.2 5x5 image and its co-occurrence matrixes with distance d=1 and with different directions ($\theta = 0^o, \theta = 45^o, \theta = 90^o, \theta = 135^o$).

The value of can be d=1,2 and $\theta = 0^0, 45^0, 90^0, 135^0$ can be for co-occurrence matrix calculation. Texture base features can be explained by using this co-occurrence matrix. Following are the most widely used features[26].

Correlation $$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{ijp(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \qquad 3.16$$

Energy $$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [p(i,j)]^2 \qquad 3.167$$

Inertia $$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-j)^2 p(i,j) \qquad 3.178$$

Absolute Value $$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}|i-j|\,p(i,j) \qquad 3.19$$

Inverse difference $$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{p(i,j)}{1+(i-j)^2} \qquad 3.20$$

Maximum Probability $$\max_{i,j} p(i,j) \qquad 3.218$$

Entropy $$H = -\left(\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} p(i,j)\log_2[p(i)]\right) \qquad 3.22$$

3.6 Classification Techniques

The techniques used for classification of images may be divided into following three main categories.

- Classification based on supervised learning
- Classification based on unsupervised learning
- Classification based on reinforcement learning

Classification is the method used for classifying the input data patterns into meaningful categories based on some specific criterion. The selection of an omnipotent classifier may require some important factors like the ones given below.

- Accuracy for classification
- Algorithmic performance
- Need for computational resources

3.7 Learning Methods Vs Non-Learning Method Based Classification

The process of assigning a pattern or object to a category based on the features provided by feature extractor is called as classification. There are many techniques which come under the learning free methods such as K-Nearest Neighbors classifiers, Minimum distance classifier, and Bayesian methods used voting schemes to find the closest objects. These methods do not use the learning process like ANN and SVM do. We have used both the learning free as well as learning methods for classification of tumor in this thesis.

3.7.1 Minimum Distance Classifier

The minimum distance classifier works on the mean vector of the image under consideration. The minimum distance classifier works well when the distance between means is large as compare to the spread of each class with respect to its mean. The minimum distance classifier works well when the distribution of data of each class is in the form of *hyper cloud* in n-dimensional pattern class. If the separation of data between two classes is in the form of spread then the results are not very good at all. The decision boundary separating two classes is given in the equation below.

$$d_i(X) - d_j(X) = 0 \qquad 3.23$$

The decision boundary between two classes can be defined by single function as

$$d_{ij}(X) = d_i(X) - d_j(X) = 0 \qquad 3.24$$

Thus if $d_{ij}(X) > 0$ we called it as class one and

If $d_{ij}(X) < 0$ we classify as class second.

Furthermore

$d_{ij}(X) = d_i(X) - d_j(X) = 0$ can be expressed as

$$d_{ij}(X) = X^T(m_i - m_j) - \frac{1}{2}(m_i - m_j)^T(m_i - m_j) = 0 \qquad 3.25$$

The surface given by above equation is the perpendicular bisector of the line segment joining m_i and m_j.

3.7.2 Bayes Classifier

This classifier is called as the probabilistic approach to recognition. The probability considerations are very important in pattern recognition because of the randomness of the data occurring in many natural processes. The bayes classifier is a quadratic classifier. The two important parameter used in the Bayes classifier are the mean and the covariance matrix of the data. In general the Bayes decision functions for Gaussian pattern classes are represented as:

$$d_j(X) = \ln P(w_j) - \frac{1}{2}\ln|C_j| - \frac{1}{2}\left[(X - m_j)^T C_j^{-1}(X - m_j)\right] \qquad 3.26$$

for j=1,2,3..............W.

The decisions functions in the above equation are quadratic functions in n-dimensional space.

Also in the above equation the mean and co-variance are used which are given below.

$$m_j = \frac{1}{N_j} \sum_{x \in w_j} X \qquad 3.27$$

Similarly Co-Variance matrix is given as:

$$C_j = \frac{1}{N_j} \sum_{x \in w_j} XX^T - m_j m_j^T \qquad 3.28$$

The Co-variance matrix given above is symmetric and positive semi definite. So for the Bayes classifier there are two parameters which are called as the mean and the covariance as compare to only mean in the minimum distance classifier.

3.7.3 Multi Support Vector Machine Based Classification

Support vector machine (SVM) is one of the method used in many application for the classification task. In this method we looks for the optimal separating hyperplane, typically in a generalized feature space of much higher dimension than the original feature space. SVM has also been applied on different real world problems such as face recognition[39], text categorization[40], cancer diagnosis[41], glaucoma diagnosis, microarray gene expression data analysis. Proposed system used SVM for binary classification of brain MR image as normal or tumor affected. SVM basically tries to divide the given data into decision surface. Decision surface is a hyperplane which divides the data into two classes. Training points are the supporting vector which defines the hyperplane. Figure below shows the simple linear support vector machine. The basic theme of SVM is to maximize the margins between two classes of the hyperplane[42][43].

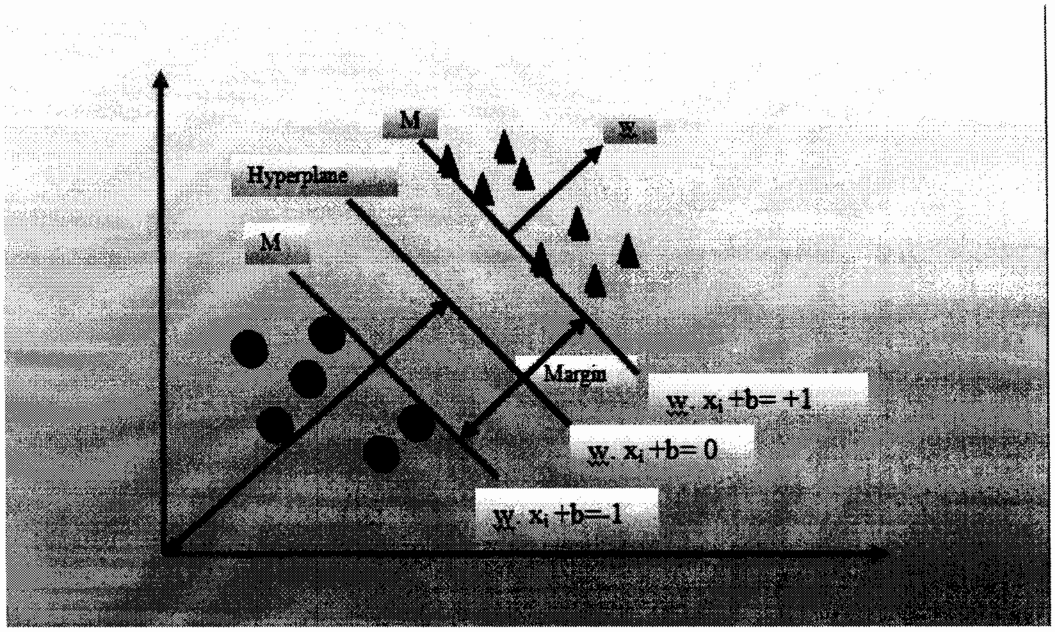


Figure 3.6 A Linear Support Vector Machine

We have used multi support vector machines instead of only one for our classification. In this method different features are given to different SVM and the results were analyzed

CHAPTER 4

PROPOSED METHODOLOGY

4.1 Introduction

Separating touching objects from an image is the most difficult image processing task. In many computer vision applications the process of separation of an image into regions of similar characteristic is called as image segmentation. So in this thesis a method for segmentation of brain tumor has been developed. From medical perspective the process of isolating a tumor from the rest of the tissues is very time consuming. Manual segmentation may take approximately 2-6 hours for the physician to isolate a tumor from other tissues in the brain. So the need arise for automatic segmentation of tumor tissues as compared to the rest of the tissues. A lot of research has been contributed in this direction and promising results have been achieved. Normally in the text on MRI, image segmentation methods can be divided in to two main branches. One is called the single image, or intensity value based, segmentation in which a single 2-D or 3-D image is used and, there are multispectral segmentation methods, in which many images with dissimilar gray scale are presented. We propose in this thesis a new technique which is called as Morphological Watershed segmentation followed by graph based techniques. So the basic theory of watershed segmentation is explained in the text given below.

4.2 Morphological Watershed Transform

Morphological watershed segmentation can be best explained by visualizing an image in three dimensions. The two axes are called the spatial coordinates and the third one is called the gray levels. Furthermore topographic and hydrology concepts can also be helpful in the explanation of watershed segmentation methods. So in this context the high amplitude pixel values corresponds to ridge lines and low amplitude pixels corresponds to valley points. Now suppose that a drop of rain water falls at any point of the surface altitude, it will go towards a lower altitude and finally to a local minimum. The gathering of the whole water in the surrounding area of local minimum is called a *catchment basin.* So, all the points that come to a general catchment basin are called the *"watershed"*. A valley is a topographic region which is enclosed by ridge lines. A *ridge* is the watershed lines at which waterfalls will ultimately goes to local minimum. There are basically two techniques to compute the watershed of an image: One is called the rainfall and the other is called as the flooding method.

These basic concepts are illustrated with the help of following figures.

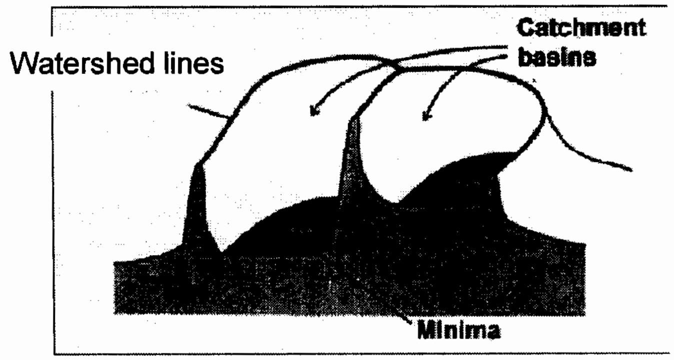


Figure 4.1Three main points of Watershed segmentation method

In the *rainfall* technique, local minima can be found from the whole image and have been assigned a distinct tag. In the next step a water drop is located at untagged pixel and allows moving towards lower amplitude value. This water drop moves down and reaches a tagged value and becomes part of the local minima.

In the *flooding* approach, we start with all the pixels having the lowest values. These values are the basis for initial watershed. For every group of intensity values if there is new values of intensity add to the adjacent regions else mark as boundary.

The main problem of watershed segmentation is over segmentation [31]. This problem arises due to noise in the input image and also due to irregularities of the gradient.

There are many researchers which have used many different methods to avoid this over segmentation problem. One way is to use morphological operators [32] to remove this over segmentation problem. Others used marker and some used graph based methods to get rid of these problems. One of the researchers [33] has used watershed base segmentation on a pyramid to solve the problem of over segmentation. The step by step procedure of the work done in this chapter is shown in the figure below.

4.2.1 Proposed Algorithm

In this method we proposed two step algorithm to overcome the problem of over segmentation in the original watershed segmentation method. We used morphological operators for internal and external markers computation [32]. The internal markers correspond to the objects of interest in the image while back ground corresponds to the back ground.

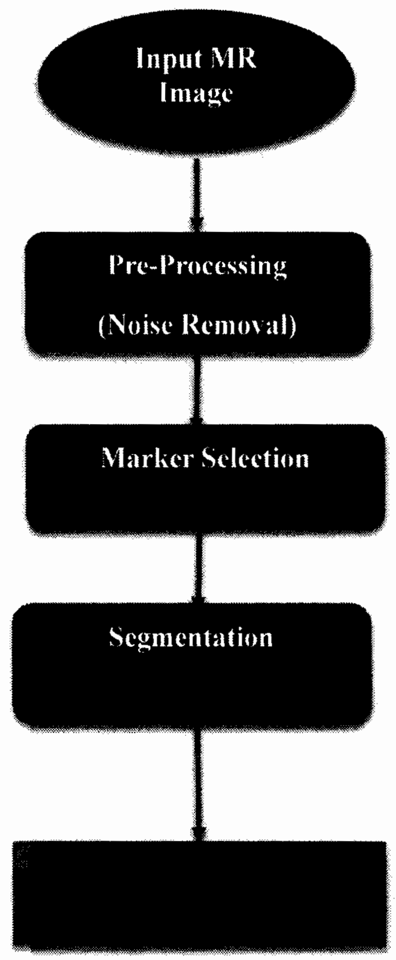


Figure 4.2 Proposed Flow chart for Segmentation

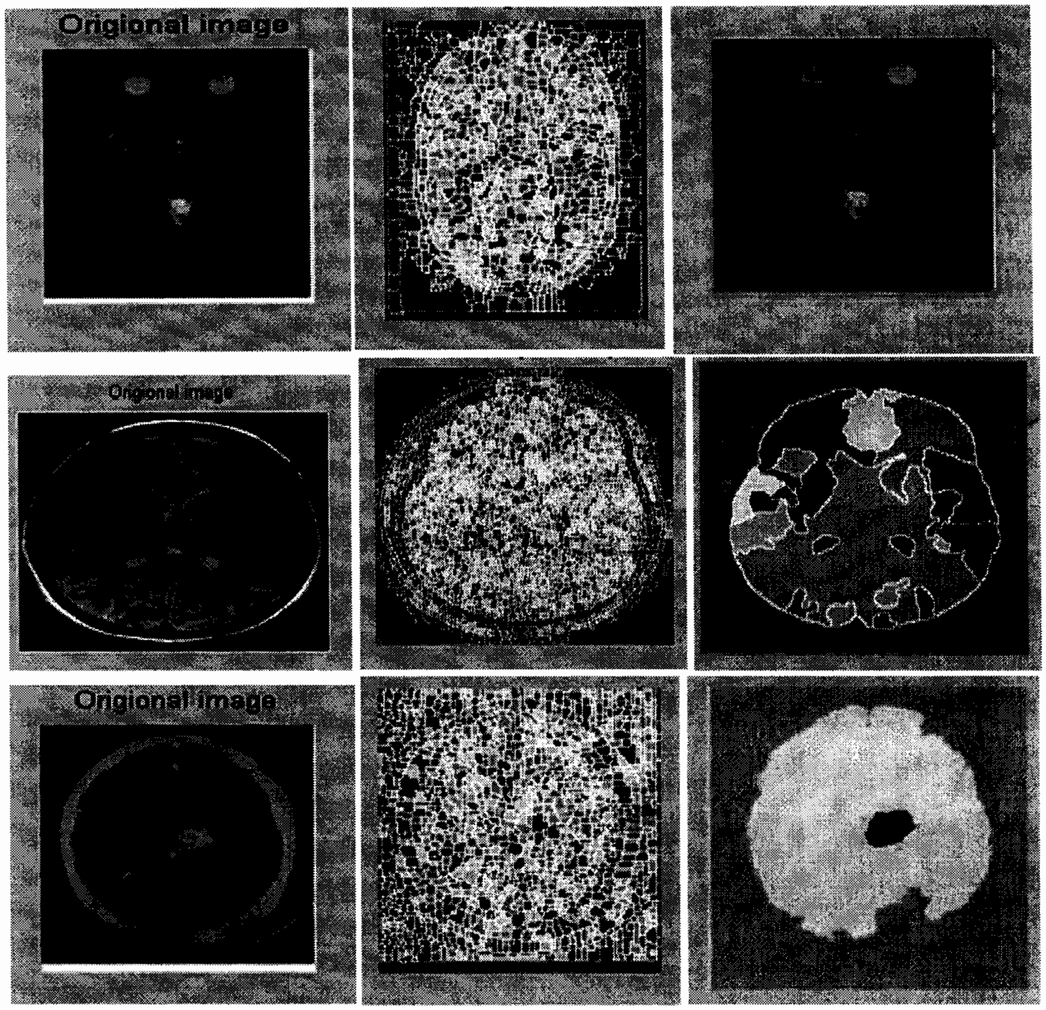
The markers for the back ground were obtained by morphological erosion of the internal markers complements. The accuracy of the algorithm is mainly dependent on the computation of internal markers. The variation in size of the biopsies made it difficult for internal marker computation. So by considerable application of morphological operators we calculated internal and external markers.

4.2.2 Simulation Results

For the evaluation of this research work 30 MRI images were used to test the proposed algorithms. In these images we have to segment the White matter, Gray matter and CSF in order to make a diagnosis. The images were obtained from the Brain Atlas web page. It was very difficult to segment the images due to high contrast, textured regions and low spatial resolution. The original images were in the color format but we converted them to gray level for easy processing.

The Algorithm was implemented in Matlab 7.9 and we used the standard function available in the tool box. The results of the proposed algorithm are shown in Fig 4.3.

To obtain the object markers or internal markers, morphological opening by reconstruction and closing by reconstruction operators were applied. The back ground markers or the external markers were obtained by eroding the complement to the internal markers. Finally the watershed transform is applied to segment the image by incorporating the previously calculated internal and external markers.



a) Original image b) Over segmentation c) segmented image

Figure 4.3 MR Image segmentation results

4.2.3 Conclusions

We have developed an automatic segmentation method which is simple, robust and gives continuous boundaries. The proposed system is divided into three main steps i.e. preprocessing, internal and external marker calculation and finally segmentation of objects from MR images. This algorithm automatically adopts to the different characteristics of the MR biopsies. Also this algorithm is very useful tool for textured images which is not possible with conventional methods. This method reduces the computational cost to be implemented in hardware. MR images have high spatial resolution and the computational cost is added to the marker selection and watershed transform. All experiments show that the proposed system gives remarkably good results as compared to the recently proposed techniques.

4.3 Automatic Brain Tumor Classification Using Neural Network

Classification is used for classifying the objects into corresponding classes. For classification of the images different features of the image are extracted. These features are used for classifying the brain MR image as benign and malignant. Features extraction and features selection from the image is very important task. Because good and effective features selection plays a vital role in the performance of the classification. Good accuracy of the classifier can be achieved by the selection of optimum feature set.

4.3.1 Proposed Method

The proposed method consists of multiple steps. Figure 4.4 shows the details of the proposed method. First the MR images for analysis are loaded in to the Matlab work space. In the second phase noise from the images are removed by using morphological operators. In the third phase discrete wavelet transform using simlet are used for feature extraction. The wavelet features were extracted up to three levels and then statistical signal processing tool was used for feature reduction by preserving maximum information. The detail of each step is given below.

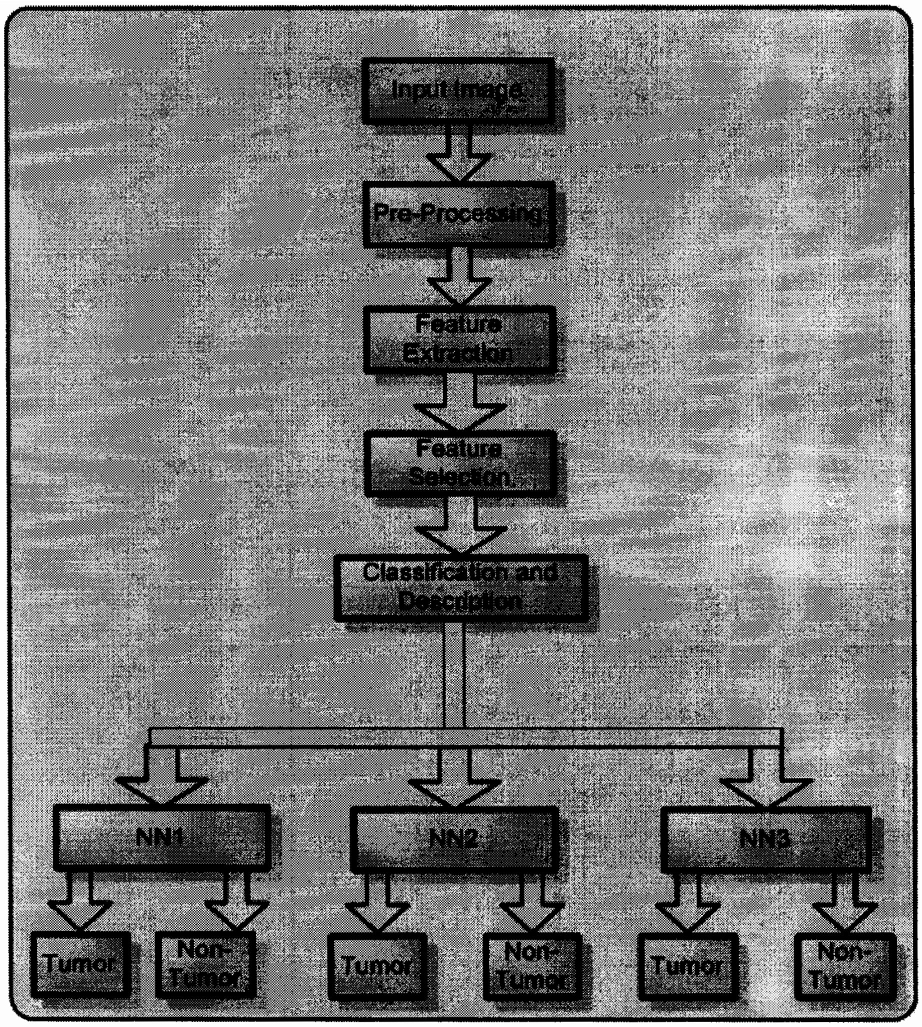


Figure 4.4 Flow Diagram of Proposed System

4.3.2 DWT Based Feature Extraction

Wavelet transform is a power tool used in the image processing these days. Discrete Fourier transforms which is composed of infinite number of sine and cosine orthogonal basis functions. On the other hand wavelet transform uses waves of limited duration are called as wavelets. These wavelets are of varying frequency and limited duration.

In the Fourier domain we only get the frequency based information within an image and do not give any information about the temporal relation. The wavelet transform is a powerful tool giving both the frequency as well as temporal information about an image.

Wavelet transform can be used for feature extraction, demising, and for compression purposes. We in this thesis using wavelet transform for feature extraction.

Wavelet transforms decomposes an image at different level of resolution. If the objects are small in dimension or low in the contrast, we study them at high resolution, if the objects are large in dimension or vary high in contrast, a coarse examination is all that is necessary. In contrast to this if the image contains both high resolution and low resolution it is useful to study them at more than a few different level of resolution. A basic decomposition using wavelet transform is illustrated in the following schematic diagram.

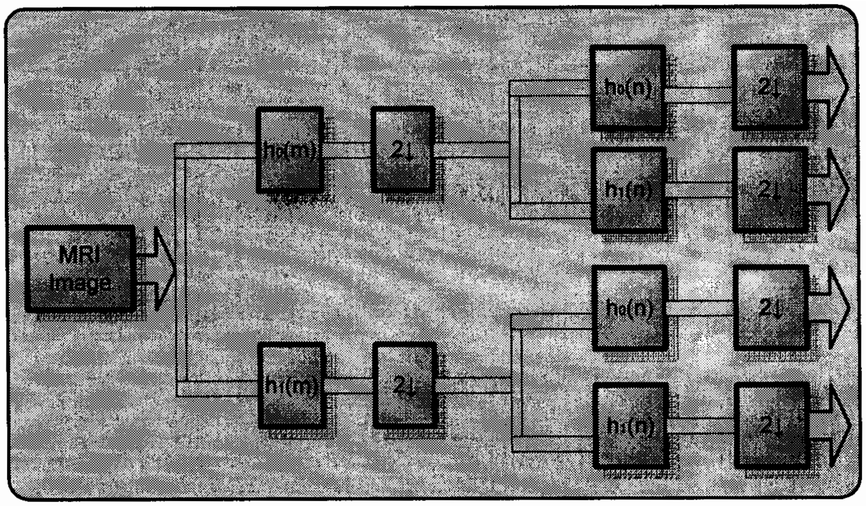


Figure 4.5 DWT decomposition diagram

4.4 PCA Based Feature Selection for Tumor Classification

The goal of feature extraction is to simply represent the original information in a lower dimensional space by retaining most of the relevant information. So this task can be accomplished by using PCA.

The features extracted by using DWT are large in numbers so computational time required is very much high when processing all of them. So there is a requirement to reduce the number of features in such a way that maximum information remains intact. In this thesis we have used Principal Component Analysis as tool for feature reduction.

PCA is a useful statistical tool that has found application in many field such as face recognition and image compression. It is also used for patterns of high dimensionality. The basic concept of PCA is to use the statistical parameters such as mean, standard deviation and covariance matrix calculation. The basic steps used in PCA are listed below.

First of all get the data of all the classes in the matrix form. Each column of data should represent the features of image under process. After this calculate the mean of the whole matrix which becomes the mean vector. The calculation of mean is described as under.

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad 4.1$$

For PCA to work properly subtract the mean from each of the data dimensions. The mean subtracted is the average across each dimension. So we get the new matrix with mean subtracted from each feature vector. This new matrix has zero mean data.

The second step is the calculation of the covariance matrix. Although in MATLAB both these steps can be accomplished by using only two commands. In the literature the formula for calculating covariance matrix is given as:

$$Cov(X) = \sum_{k=1}^{n} (X_k - m)(X_k - m)^T \qquad 4.2$$

Where m in the above equation is the mean of the data matrix and X represent input data.

For PCA the maximum information is in the variance (in case of 2-dimension data) and in the covariance for high dimensional data.

The next step is the calculation of the eigenvectors and Eigen values of the covariance matrix. The calculation of these parameters is very important because they tell us useful information about our data. For a given 2-dimensional data set the PCA components are shown below.

In general once the Eigen vectors are found from the covariance matrix the next step is order in descending order of Eigen values. This will give us the components in order of significance. After we get the larger Eigen vectors against the larger Eigen values we form a feature vector which holds maximum information.

4.5 Multi Neural Network Based Classification

We have used two methods based on linear discriminate functions. Some problems can be easily solved by using linear discriminator functions but in reality most of the problems are of non-linear nature. So we moved to Neural Network which can easily solve the Non-linear problems. In using NN the non-linearity is learned from the training data. The basic architecture of NN is shown in the figure below. In our proposed method we use more than one neural network for training and testing the available data. The basic architecture is the same in all the neural networks. Every neural network is based upon multi layers, it uses three layers, input, hidden, and output layers. Optimal features for every class have been calculated and are given as input to the neural network which contains seven neurons in the input layer.

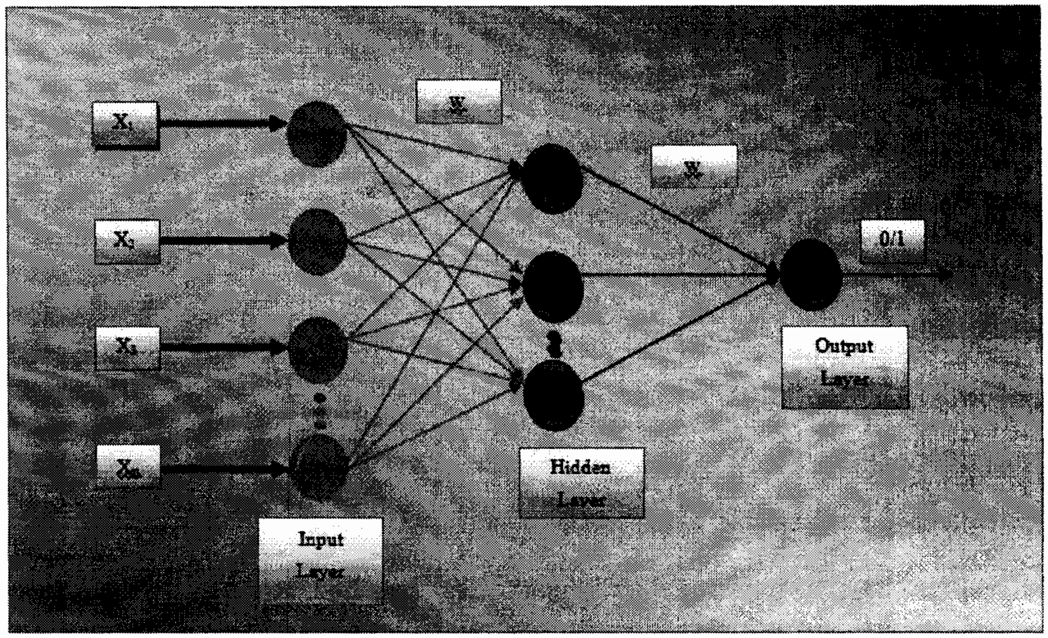


Figure 4.6 Basic Architecture of Multi layer NN

The hidden or the processing layers vary in each neural network. We have used different number of neurons in the hidden layer which comes from experience or from the guess work depending upon the optimal performance. The final layers in each network consist of only one neuron. We have used only one neuron in the output layer because we classify benign tumor from malignant ones.

Neural networks are trained on some set of images, and tested upon unseen images. We have used neural networks on the principal of *back propagation* algorithm for error computation and weight calculation for each layer.

The training of the network by *back propagatio*n involves three stages. In the first stage all the training patterns are *forwarded*, this process is called as feed forward. In the second stage which is called as back propagation, the associated error at each layer is calculated and in the third stage weights are updated based upon error calculation. This process continues in an iterative manner until the desired results have been achieved. The performance of all the neural networks is heavily dependent upon the weights, biases and transfer functions used at each layer of network architectures. There are many methods for initialization of weights of neural networks but we have randomly initialized the weight matrix at the start of algorithm.

The transfer functions used in the Feed forward back propagation neural networks are pure line (), logsig (), and tansig (). These transfer functions have some important properties like; it should be continuous, differentiable, and monotonically non-decreasing. From computational point its derivative must be easy to calculate. These transfer functions acts as a summation point and calculate the output from the inputs presented.

After the training process we verified the networks with already trained image along with some novel patterns to calculate the accuracy and error. In the testing phase only feed forward stage of back propagation neural net is utilized. In the testing phase we used the already computed optimum weights. In testing, like training the input images are required to go through the same process as feature extraction and optimum feature selection. After these optimum features have been selected are applied to the network for simulation purposes. Simulation is the process in which the data is presented and it simulates the network. After each image is presented to the network from database corresponding counter value is incremented as correct counter or error counter.

4.6 Experimental Results and Discussion

The proposed system is implemented by using Matlab 7.6 environment. The proposed technique is implemented and tested on the dataset available at [38] and also on the real brain MRI dataset available at [37]. The dataset are loaded into the Matlab work space followed by noise removal and feature extraction. After feature extraction PCA is applied for optimal feature selection and finally Neural Network is applied for classification. Performance of the classifier is measured in terms of sensitivity, specificity and accuracy. We tested the accuracy of the classifier on two different dataset shown in Table 4.1. The accuracy of the classifier is tested by using different number of wavelet coefficients. We analyzed that the accuracy is very much lower when we used 2, 3, 4, and 5 as coefficients. We tested the accuracy by using more coefficients and the results are good enough when we used 7 as the number of features. Using more than 7 features did not improve the results of the classifier which gives us an intuitive idea that these top 7 coefficients hold maximum information. The results are shown in table 4.1 and figure 4.7 respectively.

Methods used	AANLIB dataset Accuracy (%)	Real MRI Data Accuracy (%)
DWT+ ANN	95.30	94.55
DWT+MANN	97.13	96.72
DWT+MSVM	97.25	96.64

Table 4.1 Comparison of the classifier performance with other classifiers

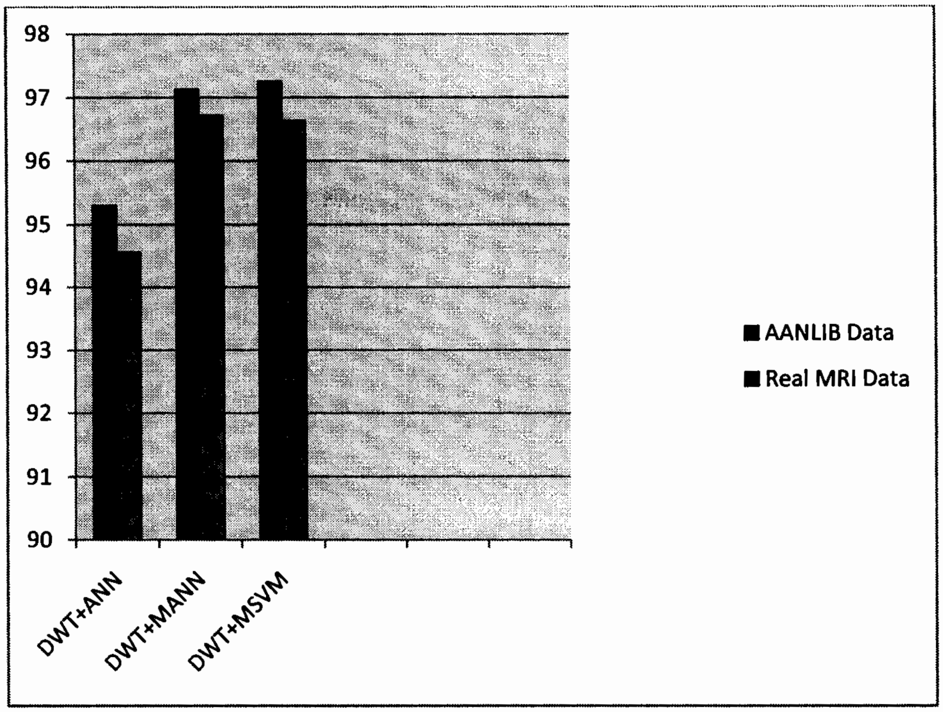


Figure 4.7 Results of Table 4.1 by Bar Graph

	Accuracy %	Specificity %	Sensitivity %	Error Rate %
Real MRI Dataset[37]	99.89	99.45	99.75	2.4
AANLIB[38] Dataset	99.64	93.65	94.93	3.1

Table 4.2 Performance Measures for Available Dataset

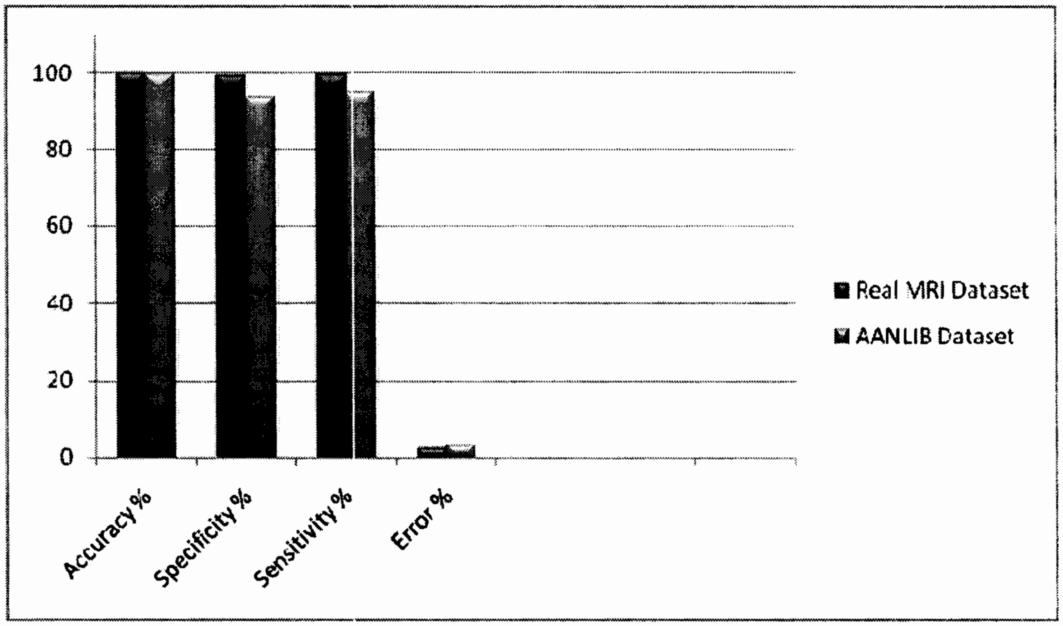


Figure 4.8 Results of Error! Reference source not found. **shown by the bar graph**

Table 4.2 and figure 4.8 shows the results of our classifier in terms of quantitative performance metrics. Specificity and sensitivity rates for both the dataset are good enough and the error rate is very low.

4.7 Conclusion

We have proposed an automated system for brain tumor classification in this last part of the current chapter. The proposed system is divided in to multiple phases in term of feature extraction, feature reduction and finally classification. The performance of the proposed system is good as compared to the recently proposed techniques.

CHAPTER 5

CONCLUDING DISCUSSION

5.1 Conclusion

In this dissertation various methods were applied for effective tumor detection and classification. Manual Brain tumor extraction for physicians is very time consuming and tedious task. Moreover unavailability of expert physicians round the clock is a big challenge as it is very difficult to afford them. In this dissertation computer diagnosis method was developed with the intention to help out radiologist and other physicians. This method is noninvasive and only runs on a machine so it is not as much costly. This problem has been perceived as image processing so various methods and techniques have been applied to give its solution. The different steps involved in this dissertation are discussed as under.

The images given to the system needs to be classified a normal or Tumorous before segmentation. Segmentation is very important step for tumor detection. Morphological watershed transform has been applied followed by graph based technique for tumor detection. Feature extraction has been done by using multi resolution discrete wavelet transform. After feature extraction classification is all that required for final discrimination.

Classification is the procedure in which different objects according to their properties are classified into different classes. So for this purposes neural network and support vector machine are used as learning based classification methods. Also a novel approach called as Multi neural network and multi support vector machine techniques have been used for tumor classification in this thesis. These techniques are called as quantitative performance analysis methods. The other type of classification called as assignment classification has also been used as classifier.

5.2 Future Directions

We have used different method in this dissertation for detecting tumor and classification. Our focus was only to detect benign tumor in comparison with malignant ones. The next study can be the classification of sub type of malignant (aggressiveness) tumors. The other tumors types such as meningiomas, medulloblastomas, schwannomas, and germ cell tumors.

Furthermore classification of these subtypes of tumors can be done on grade systems. Grade of the tumor refers to its cell shape. This grade system may start from (grade 1) low grade to higher grade (grade 4). Cells from higher grade are considered more abnormal as compared to low grade tumor.

In future other classification methods such fuzzy c-means, fuzzy curvelet transform, k-means and ensemble based classifiers can be used for further enhancing the accuracy of classifiers. In terms of features, local features like SIFT textured and fractal dimensions can be used for best feature extraction.

REFERENCES

[1] "www.wikipedia.com."

[2] "http://huehueteotl.wordpress.com/2008/10/."

[3] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision,* vol. 60, 2004, pp. 91–110.

[4] R.M. Haralick and L.G. Shapiro, "Image segmentation techniques," *Computer vision, graphics, and image processing,* vol. 29, 1985, pp. 100–132.

[5] R.C. Gonzalez and R.E. Woods, *Digital image processing. 2008*, Pearson Prentice Hall, Upper Saddle River, NJ, .

[6] Y. Bazi, L. Bruzzone, and F. Melgani, "Image thresholding based on the EM algorithm and the generalized Gaussian distribution," *Pattern Recognition,* vol. 40, 2007, pp. 619–634.

[7] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic imaging,* vol. 13, 2004, pp. 146–168.

[8] L. Djerou, N. Khelil, H.E. Dehimi, and M. Batouche, "Automatic Multilevel Thresholding Using Binary Particle Swarm Optimization for Image

Segmentation," *2009 International Conference of Soft Computing and Pattern Recognition*, 2009, pp. 66–71.

[9] J.S. Weszka and A. Rosenfeld, "Threshold evaluation techniques," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 8, 2007, pp. 622–629.

[10] Y. Chen and J.Z. Wang, "A region-based fuzzy feature matching approach to content-based image retrieval," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, 2002, pp. 1252–1267.

[11] A. Ghosh and S.K. Pal, "Neural network, self-organization and object extraction," *Pattern Recognition Letters*, vol. 13, 1992, pp. 387–397.

[12] "http://www.advancedimaging.info/mri/."

[13] "http://archive.student.bmj.com/issues/03/11/education/402.php."

[14] Y. Xu and F. Song, "Feature extraction based on a linear separability criterion," *International Journal of Innovative Computing, Information, and Control. v4 i4*, pp. 857–865.

[15] G.J. Mun, B.N. Noh, and Y.M. Kim, "ENHANCED STOCHASTIC LEARNING FOR FEATURE SELECTION IN INTRUSION CLASSIFICATION."

[16] I.G.P. Wijaya, K. Uchimura, and Z. Hu, "Face Recognition Based on Dominant Frequency Features and Multiresolution Metric," *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*, 2008, p. 9.

[17] A. Ohno and H. Murao, "A Similarity Measuring Method for Images Based on the Feature Extraction Algorithm using Reference Vectors," *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*, 2008, p. 454.

[18] R.M. Haralick, K. Shanmugam, and I.H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man and cybernetics*, vol. 3, 1973, pp. 610–621.

[19] H.K. Ekenel and R. Stiefelhagen, "Local appearance based face recognition using discrete cosine transform," *13th European Signal Processing Conference (EUSIPCO 2005), Antalya, Turkey*, 2005.

[20] S.A. Khayam, "Syed Ali Khayam," *Signal Processing: Image Communication*, vol. 18, 2003, pp. 575–595.

[21] G. Strang, "The discrete cosine transform," *SIAM review*, vol. 41, 1999, pp. 135–147.

[22] G. Sun, X. Dong, and G. Xu, "Tumor tissue identification based on gene expression data using DWT feature extraction and PNN classifier," *Neurocomputing*, vol. 69, 2006, pp. 387–402.

[23] S. Li, C. Liao, and J.T. Kwok, "Wavelet-based feature extraction for microarray data classification," *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, 2006, pp. 5028–5033.

[24] M. Kociołek, A. Materka, M. Strzelecki, and P. Szczypiński, "Discrete Wavelet transform–Derived features for digital Image texture Analysis," *Proc. of Interational Conference on Signals and Electronic Systems*, pp. 163–168.

[25] O.S. Al-Kadi and D. Watson, "Susceptibility of texture measures to noise: an application to lung tumor CT images," *BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on*, 2008, pp. 1–4.

[26] A. Materka and M. Strzelecki, *Texture Analysis Methods–A Review, Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels 1998.*

[27] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern classification*, Citeseer, 2001.

[28] S. Rajasekaran and G.A. Pai, *Neural networks, fuzzy logic and genetic algorithms: synthesis and applications*, PHI Learning Pvt. Ltd., 2004.

[29] S. Haykin, *Neural networks: a comprehensive foundation*, Prentice Hall PTR Upper Saddle River, NJ, USA, 1994.

[30] W.K. Pratt and J. Wiley, *Digital image processing: PIKS inside*, Wiley, 2001.

[31] A.M. López, F. Lumbreras, J. Serrat, and J.J. Villanueva, "Evaluation of methods for ridge and valley detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 21, 2002, pp. 327–335.

[32] L. Najman and M. Schmitt, "Geodesic saliency of watershed contours and hierarchical segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 18, 2002, pp. 1163–1173.

[33] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 13, 2002, pp. 583–598.

[34] H. Duda and P. Hart, *Stork, Pattern Classification*, John Wiley & Sons, 2001.

[35] V. Bruce, P.R. Green, and M.A. Georgeson, *Visual perception: Physiology, psychology, & ecology*, Psychology Pr, 2003.

[36] R. Bryll, R. Gutierrez-Osuna, and F. Quek, "Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets," *Pattern Recognition,* vol. 36, 2003, pp. 1291–1302.

[37] "http://www.abrarmrict.com/."

[38] "http://med.harvard.edu/AANLIB/."

[39] J. Huang, V. Blanz, and B. Heisele, "Face recognition using component-based SVM classification and morphable models," *Pattern Recognition with Support Vector Machines*, 2002, pp. 531–540.

[40] T. Joachims, “Text categorization with support vector machines: Learning with many relevant features,” *Machine Learning: ECML-98*, 1998, pp. 137–142.

[41] T.S. Furey, N. Cristianini, N. Duffy, D.W. Bednarski, M. Schummer, and D. Haussler, “Support vector machine classification and validation of cancer tissue samples using microarray expression data,” *Bioinformatics*, vol. 16, 2000, p. 906.

[42] S.R. Gunn, “Support vector machines for classification and regression,” *ISIS technical report*, vol. 14, 1998.

[43] B. Sch\ölkopf and A.J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*, the MIT Press, 2002.