CHAPTER 1

INTRODUCTION

Brain tumor is one of the most lethal human diseases, with the need for novel therapies remaining an unmet clinical challenge. A survey on statistics of brain tumor revealed that each year more than 2,00,000 people in the United States are diagnosed with a primary or metastatic brain tumor [1]. In India, the International Agency for Research on Cancer estimated indirectly that about 635 000 people died from cancer in 2008, representing about 8% of all estimated global cancer deaths and about 6% of all deaths in India [2]. Data regarding frequencies of various primary brain tumors (diagnosed according to the World Health Organization (WHO) classification), in 3936 pediatric patients (<18 yrs of age), was collected from seven tertiary care hospitals in India. Brain tumors are the leading cause of solid tumor cancer death in children under the age of 20. They are the second leading cause of cancer death in male adults ages 20-29 and the fifth leading cause of cancer death in female adults ages 20-39. Metastatic brain tumors, cancer that spreads from other parts of the body to the brain, are the most common types of brain tumors. They occur in 10-15% of people with cancer. Primary brain tumors generally do not metastasize to other parts of the body.

There are currently no known causes of brain tumors, however, epidemiological studies are ongoing. Complete and accurate data on all primary brain tumors is needed to provide the foundation for investigations of its causes and research leading to improved diagnosis and treatment. Brain tumors have no socio-economic boundaries and do not discriminate among gender or ethnicity. At this time, brain tumor research is underfunded and the public remains unaware of the magnitude of this disease. The cure rate for most brain tumors is significantly lower than that for many other types of cancer. Males have a 0.66% lifetime risk of being diagnosed with a primary malignant brain tumor and a 0.50% chance of dying from a brain tumor. Females have a 0.54% lifetime risk of being diagnosed with a primary malignant brain tumor and a 0.41% chance of dying from a brain tumor.

There are over 120 different types of brain tumors, making effective treatment very complicated. Because brain tumors are located at the control center for thought, emotion and movement, their

effects on an individual's physical and cognitive abilities can be devastating. At present, brain tumors are treated by surgery, radiation therapy, and chemotherapy, used either individually or in combination. No two brain tumors are alike. Prognosis, or expected outcome, is dependent on several factors including the type of tumor, location, response to treatment, an individual's age, and overall health status. An estimated 35% of adults living with a primary malignant brain or CNS tumor will live five years or longer. Brain tumors in children are different from those in adults and are often treated differently. Although over 72% percent of children with brain tumors will survive, they are often left with long-term side effects.

In the recent century, increasingly use of sophisticated laboratory tests has made diagnosis a sensitive and accurate issue. The use of computer technology in medical decision support is now widespread and pervasive across a wide range of medical area, such as brain tumor research, gastroenterology etc. To identify a tumor, a patient will undergo several tests. Most commonly Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are used to locate brain tumor. The information obtained will influence the treatment a patient will receive. Perhaps the most widely used clinical diagnostic and research technique is MRI. It's an efficient medical imagery tool that has different methods (T1,T2, ARM, …) having each particular property and an effective way that enables to clarify the various tissues and to obtain a 2D, 3D and even 4D sight (3D+T) of a part of the body, in particular of the brain. It's based on the principal of nuclear magnetic resonance (NMR).

Actually, many medical imagery diagnosis systems have to face the problem of cells and their nuclei separation from the rest of the image content. As the process of separation is very important, much attention in the construction of the expert diagnosis system has to be paid to the segmentation & features extraction stage. In studying human brain, magnetic resonance imaging (MRI) plays an important role in progressive researches. Magnetic resonance (MR) imaging was introduced into clinical medicine and has ever since assumed an unparalleled role of importance in brain imaging. Magnetic resonance imaging is an advanced medical imaging technique that has proven to be an effective tool in the study of the human brain. The rich information that MR images provide about the soft tissue anatomy has dramatically improved the quality of brain pathology diagnosis and treatment.

1.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI), nuclear magnetic resonance imaging (NMRI), or magnetic resonance tomography (MRT) is a medical imaging technique used in radiology to visualize internal structures of the body in detail. MRI makes use of the property of nuclear magnetic resonance (NMR) to image nuclei of atoms inside the body. MRI provides good contrast between the different soft tissues of the body, which makes it especially useful in imaging the brain, muscles, the heart, and cancers compared with other medical imaging techniques such as computed tomography (CT) or X-rays.

Magnetic resonance (MR) imaging is currently an indispensable diagnostic imaging technique in the study of the human brain [3]. It is now one of the fastest-growing areas of medical technology. The modalities usually used to obtained medical images are X-rays, Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and ultrasound imaging. In medical imaging, one of the primary diagnostic and treatment evaluation tools for brain interpretation has been magnetic resonance imaging (MRI). It has been a widely-used method of high quality medical imaging, especially in brain imaging where MR's soft tissue contrast and non invasiveness are clear advantages. MR images can also be used to determine normal and abnormal types of brain. Moreover, the MRI characteristics will help the doctor to avoid the human error in manual interpretation of medical content. MRI is one of the scanning devices which uses magnetic fields to capture images onto films. Due to its outstanding soft tissue contrast and detailed resolution, MRI is used in anatomical assessment of human brain structures.

The MRI machine is a large, cylindrical (tube-shaped) machine that creates a strong magnetic field around the patient. The magnetic field, along with a radiofrequency, alters the hydrogen atoms' natural alignment in the body. Computers are then used to form a two-dimensional (2D) image of a body structure or organ based on the activity of the hydrogen atoms. Cross-sectional views can be obtained to reveal further details. MRI does not use radiation, as do x-rays or computed tomography (CT scans).

A magnetic field is created and pulses of radio waves are sent from a scanner. The radio waves knock the nuclei of the atoms in your body out of their normal position. As the nuclei realign back into proper position, they send out radio signals. These signals are received by a computer that analyzes and converts them into an image of the part of the body being examined. This image appears on a viewing monitor. Some MRI machines look like narrow tunnels, while others are more open.

Magnetic resonance imaging (MRI) may be used instead of computed tomography (CT) in situations where organs or soft tissue are being studied, because bones do not obscure the images of organs and soft tissues, as they do in CT. Because radiation is not used, there is no risk of exposure to radiation during an MRI procedure. Due to the use of the strong magnet, MRI cannot be performed on patients with implanted pacemakers, intracranial aneurysm clips, cochlear implants, certain prosthetic devices, implanted drug infusion pumps, neurostimulators, bone-growth stimulators, certain intrauterine contraceptive devices, or any other type of iron-based metal implants. MRI is also contraindicated in the presence of internal metallic objects such as bullets or shrapnel, as well as surgical clips, pins, plates, screws, metal sutures, or wire mesh.

Figure 1.1: MRI machine

The scanning of brain in MRI machine produces images in the form of slices as shown below

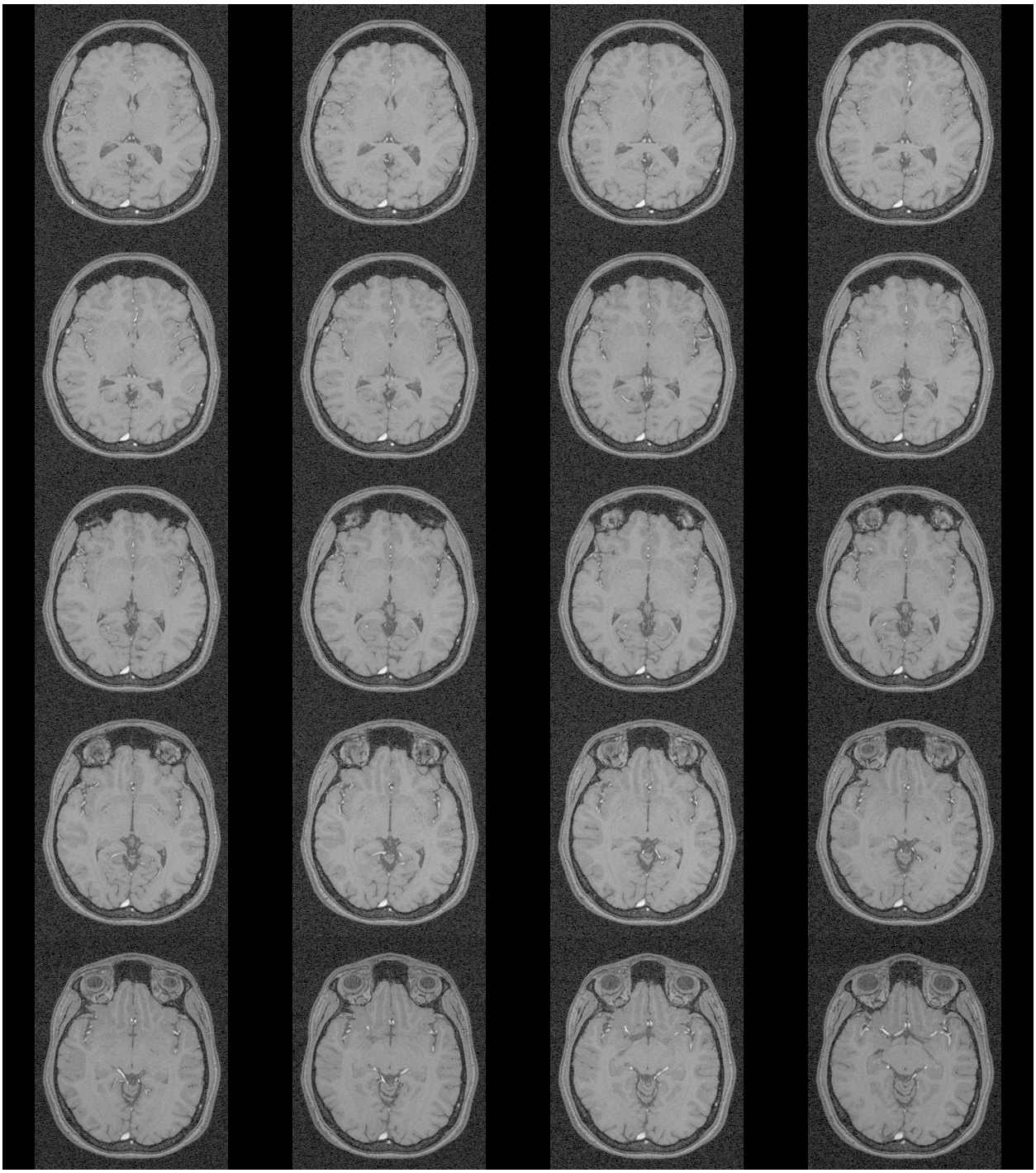
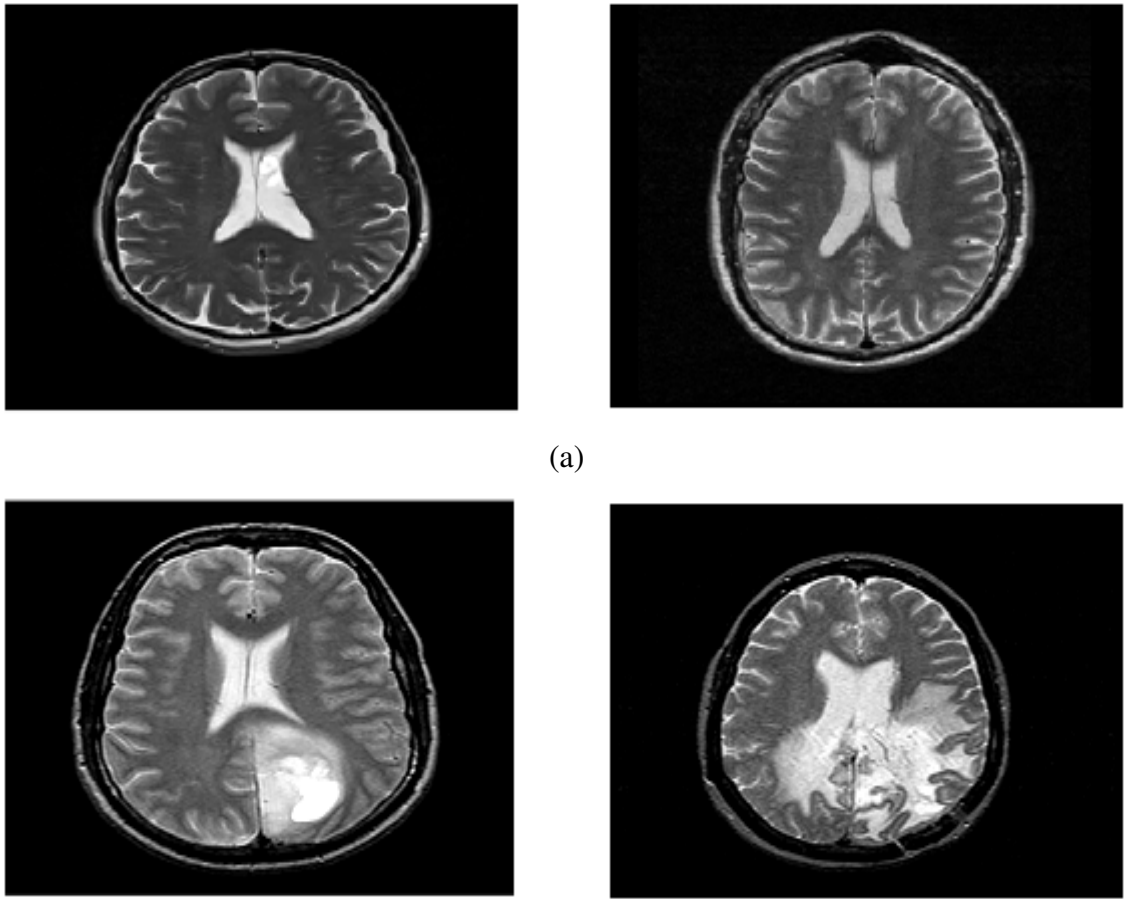


Figure 1.2: A sequence of MR image slices

The brain MRI images may contain both normal and defective abnormal slices. Normal and abnormal brain image are determined by its symmetry at the axial and coronal images. Asymmetry beyond a certain degree is a sure indication of the diseased brain and this has been exploited for initial classification at a gross level. Therefore, further examination involving MRI brain classification on the images is required.



(a)

(b)

Figure 1.3: (a) Normal MRI images (b) Abnormal MRI images

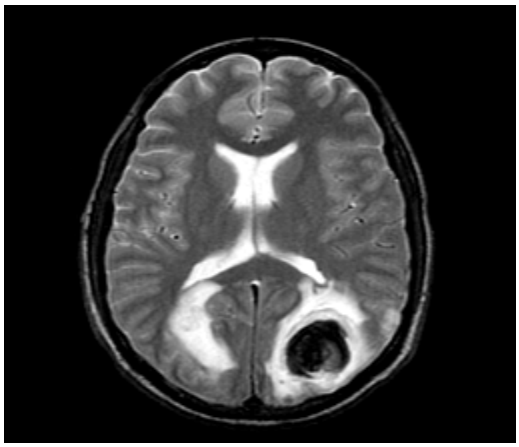


Figure 1.4: Brain Tumor Infected M.R.I Image

The most important advantage of MR imaging is that it is a non-invasive technique. The use of computer technology in medical decision support is now widespread and pervasive across a wide range of medical area, such as cancer research, gastroenterology, heart diseases, brain tumors, etc. Fully automatic normal and a pathological brain, suffering from brain lesion classification can be obtained from magnetic resonance images, which is a great importance for research and clinical studies. The rich information that MR images provide about the soft tissue anatomy has dramatically improved the quality of brain pathology diagnosis and treatment. However, the amount of data is far too much for manual interpretation and hence there is a great need for automated image analysis tools. Pattern recognition techniques are being increasingly employed in magnetic resonance imaging (MRI) data analysis.

1.1.1 MRI Analysis using Image Processing

The Images obtained using MRI scanning are used in Machine intelligence for detection of diseases like brain tumor using image processing techniques. For this algorithms are to be developed so that the normal & abnormal MRI Images can be classified by machine or computer.

The MRI Image undergoes series of following steps for analysis using image processing techniques.

A). Image Segmentation & Features extraction:

Firstly the MRI Image is segmented & features are extracted using various image processing techniques. Various schemes used for Image Segmentation and features extraction are:

1) Discrete Wavelet Transformation

2) Edge based segmentation techniques include:

a) Roberts & Prewitt filters
b) Sobel filters
c) Canny Edge filters
d) Laplacian filters

Features extraction schemes that can be used in above Edge based segmentation techniques are including Mahalanobis distances, Euclidian distance etc.

3) Region based segmentation techniques like Region Growing, Watershed algorithm, Thresholding etc.Features extraction schemes that can be used in above Region based segmentation techniques are including Texture features or more precisely, GrayLevel Co-occurrence Matrix (GLCM) features, Mahalanobis distance, Euclidian distance etc.

B). Features Reduction:

After Features extraction the dominant features are selected using Principal component analysis.

C). Classification:

After dominant features vectors are selected, a classifier is to be selected for training & classification. Various schemes of classifiers are available as:

a) Artificial Neural Network (A.N.N)
b) Maximum Likelihood
c) k-Nearest Neighbors (k-NN)
d) Parzen Window Method

1.2 Previous Work

The automatic classification of brain magnetic resonance images is vital process for separating healthy subjects and subjects with various brain diseases like, cerebrovascular, Alzheimer, brain tumor, inflammatory, etc. The process of automatically classifying MR image is a challenging process. This leads to many different approaches. Recent research work has shown that features extraction of human brain in magnetic resonance (MR) images is possible via various methods like Region based segmentation techniques like Region Growing, Watershed algorithm [4], Edge based segmentation techniques. Wavelet Transform is best suitable method for features extraction among all above methods according to latest research papers. So, in our Thesis work wavelet transform is used as features extraction method of Brain MRI Images. For features reduction principal component analysis is best suitable method used widely in various latest research papers. For classification a large variety of methods are discussed in the literature. Sonka and Fitzpatrick in [5] provide a review of classification methods ranging from computer vision through statistical approaches to machine learning. Besides this various unsupervised classification techniques such as self-organization map (SOM) and fuzzy *c*-means [6] & other supervised classification techniques, such as Maximum Likelihood, Parzen Window Method are widely used in latest research to classify the normal/pathological T1 and T2-weighted MRI images.

1.3 Thesis Objective

In this thesis, our goal is to achieve higher classification rate in order to diagnosis normal images from those with brain abnormality like Brain Tumor. Firstly, features are extracted using discrete wavelet transformation. Wavelets seem to be a suitable tool for this task, because they allow analysis of images at various levels of resolution. Secondly, principal component analysis (PCA) is used for reducing the feature vector dimension and also increasing discrimination between classes. Principal component analysis is appealing since it effectively reduces the dimensionality of the data and therefore reduces the computational cost of analyzing new data. Finally, pattern recognition method ANN is used for classification. The results indicate classification of data. In our work the database contains different images. We use artificial neural network as a different classifier technique. The number of features obtained by PCA for maximum classification rate is less and we obtain better classification rate. However, DWT, PCA, and classifiers are commonly used steps in pattern recognition problems.

1.4 Thesis Outline

The Rest of the thesis is outlined in this section. In second chapter the basics of MRI is discussed. The physics behind the neuro imaging or MRI imaging has been discussed. The overview of MRI Imaging sessions is also discussed in this chapter. In third chapter the various methods used in the processing of MRI image are discussed. For Features Extraction wavelet Transform is discussed. Features Reduction is done using Principal Component Analysis & for classification Artificial Neural Network is discussed. The simulation results in MATLAB are discussed in chapter four. Finally, thesis ends with the conclusion and the scope for the future work in chapter five.

CHAPTER 2

THE BASICS OF MRI

Magnetic Resonance Imaging, or MRI, typically measures the response of hydrogen molecules to a perturbation while in a magnetic field. There are three basic steps involved in measuring the MRI signal. The first step is to place the brain in a magnetic field. The second step is the application of a brief radiofrequency (RF) pulse and the third step is measuring the relaxation [7]. Each of these steps will in turn be considered in the following paragraphs.

2.1 The brain in a magnetic field

The first step to measuring an MRI signal is to place the brain in a magnetic field. This will cause the atomic nuclei to align with the magnetic field. This occurs to all nuclei that are electrically charged and spin around their axis. Of the many types of nuclei in the brain, it is the hydrogen nucleus that is most commonly measured in MRI [8] [9]. This is because hydrogen nuclei are abundant in the human brain and give a strong MRI signal. Hydrogen nuclei are positively charged particles that spin around their axis. When an electrically charged particle moves, it produces a magnetic field. This magnetic field can be represented as a vector (a mathematical entity with both amplitude and a direction).

Generally, a vector is mathematically depicted as an arrow where the length of the arrow represents the amplitude of the vector and the direction in which the arrow is pointing reflects the direction of the vector. Since each hydrogen nucleus produces a magnetic field, which can be represented as a vector, this equals saying that each hydrogen nucleus in the brain can be seen as a vector with the vector representing the strength and direction of the magnetic field of the hydrogen nucleus produced by its spinning around its axis. This vector is also known as the Magnetic Dipole Moment (MDM) [7] [8] [9] .Before the brain is placed in a magnetic field, the MDM.s of each hydrogen nucleus points in a random direction: the nuclei are not aligned. When the brain is placed in a magnetic field, two things happen simultaneously [8] [9].

Firstly, the MDM.s of many of the hydrogen nuclei aligns themselves in the direction of the main magnetic field. How many of the MDM.s align themselves in the direction of the main magnetic field depends on the strength of this magnetic field. The stronger the magnetic field, the higher the percentage of the MDM.s that align themselves to the magnetic field [8] [9].

Secondly, when the brain is placed in a magnetic field the MDM.s of the hydrogen nuclei start to precess. The frequency of this precession depends first of all on the type of nucleus. This means that the MDM of a hydrogen nucleus will have a different frequency of precession from, for instance, the MDM of sodium nuclei in a certain magnetic field. Second of all, the frequency of precession depends on the strength of the magnetic field.

The frequency of precession is directly proportional to the strength of the magnetic field, so the stronger the magnetic field the higher the frequency of precession. For example, in a magnetic field of 1.5 Tesla the frequency of precession for the MDM.s of hydrogen nuclei will be 64 MHz (64000000 revolutions per second) and in a magnetic field of 3 Tesla the frequency of precession will be 128 MHz [7] .

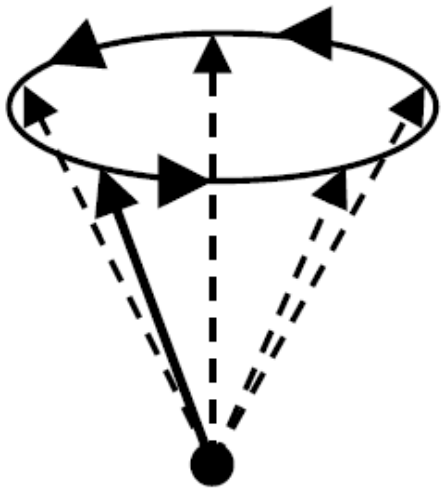


Figure 2.1: Precession of the MDM

2.2 Application of the radiofrequency pulse

The second step when measuring the MRI signal is the brief application of the radiofrequency pulse. The Radiofrequency (RF) pulse is typically an electromagnetic wave resulting from the brief application of an alternating current perpendicular to the direction of the main magnetic field, otherwise known as a 90^0 RF-pulse. The ultimate goal of this 90^0 RF-pulse is to tip the MDM.s of the hydrogen nuclei. Conventionally, the direction along the main magnetic field is referred to as the z-axis.

The 90^0 RF-pulse then basically tips the MDM.s in the x-y plane. This will only work if the frequency of the RF- pulse equals the frequency of the precession of the MDM.s. Because the MDM.s of the hydrogen nuclei have their own specific frequency of precession in a given magnetic field, it is possible to selectively tip the MDM.s of the hydrogen nuclei [8] [9].

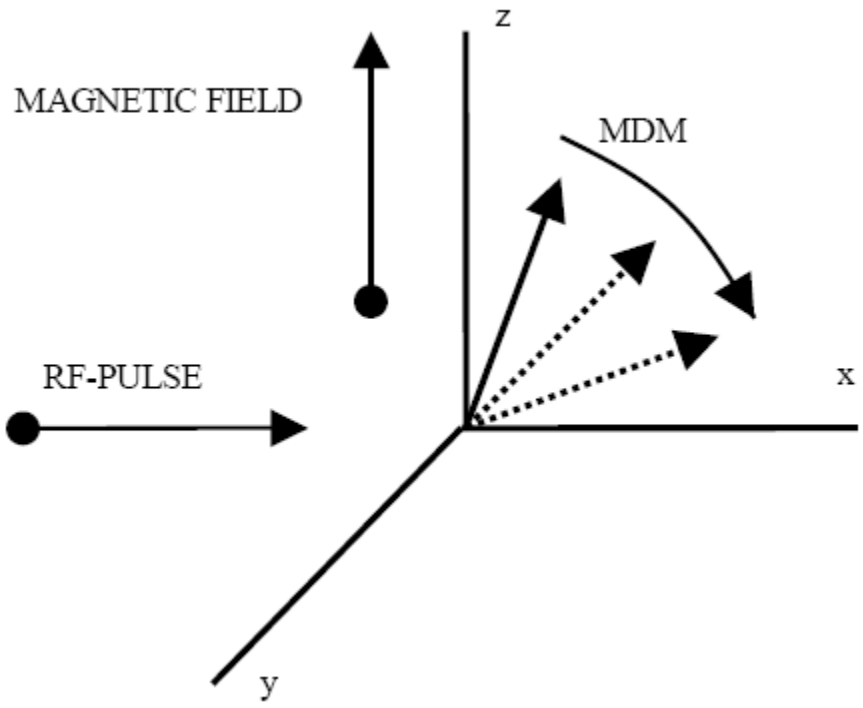


Figure 2.2: Tipping of an MDM into the x-y plane during application of the RF-pulse

After the MDM.s of the hydrogen nuclei are tipped the 90^0 RF-pulse is terminated and the MDM.s return to their original orientation. This returning to the original orientation is known as relaxation [8] [9].

2.3 Relaxation

After the 90^0 RF-pulse is terminated, the MDM.s of the hydrogen nuclei will return from their tipped state to their original lower energy state of being aligned in the direction of the magnetic field [8] [9]. Basically, the RF-pulse poured energy into the system and this energy is released when the MDM.s return to their original state. This release of energy is known as relaxation and is the signal that is measured during MRI [8] [9].

The MDM of a hydrogen nucleus can be broken down into two components. One component of the MDM is the amplitude in the z-axis. The other component of the MDM is the amplitude in the x-y plane [8] [9]. Before application of the RF-pulse the amplitude in the z-axis is maximal while the amplitude in the x-y plane is zero. Just after application of the RF-pulse the amplitude in the z-axis is zero while the amplitude in the x-y plane is maximal.

During relaxation the amplitude in the z-axis will slowly increase while the amplitude in the x-y plane slowly decreases. Therefore, the relaxation of the MDM.s of the hydrogen nuclei has two components; firstly, a re-growth along the z-axis and secondly, a decay in the x-y plane. The re-growth along the z-axis of the MDM.s is referred to as T1 relaxation. The decay in the x-y plane of the MDM.s is referred to as T2 relaxation [8] [9].

2.4 When it all comes together

The application of the 90^0 RF-pulse and the measuring of the energy released during relaxation is repeated over a vast amount of times in a typical MRI experiment. Different tissues in the brain have different T1 and T2 relaxation rates [8] [9]. This means that at each moment after termination of the RF-pulse, the amplitude of the MDM.s of the hydrogen nuclei in the z-axis and the amplitude of the MDM.s in the x-y plane will be different for different tissues.

If now the MRI signal is measured at a point after termination of the RF-pulse where either the relative difference between the amplitudes of the MDM.s of the hydrogen nuclei of different tissues in the z-axis is maximized or the relative difference between the amplitudes of the MDM.s of different tissues in the x-y plane is maximized a maximum contrast between different tissues will be obtained [8] [9]. At first, it sounds a bit counterintuitive that not the absolute difference in amplitude is maximized but instead the relative difference between the amplitudes is maximized.

When the MRI signal is measured at a point when the ratio of the amplitudes of MDM.s of different tissues in the z-axis is maximized, the signal is known as a T1 weighted signal. Alternatively, when the MRI signal is measured at a point when the ratio of the amplitudes of the MDM.s of different tissues in the x-y plane is maximized, the signal is known as a T2 weighted signal [8] [9].

By changing certain scanner parameters either a T1 weighted signal or a T2 weighted signal can be acquired [8] [9]. When the time from RF-pulse to measurement of the signal (TE) is kept short, while at the same time the time between two successive RF pulses (TR) is also kept short, the difference in T1 for the different tissues is maximized and the acquired scan is called a T1 weighted scan. T1 weighted scans are also known as anatomical scans, because they particularly show good contrast between grey and white matter.

On the other hand, when the TE is long while at the same time the TR is also long, the difference in T2 for the different tissues is maximized and the acquired scan is called a T2 weighted scan. T2 weighted scans are also known as pathological scans, because lesions appear very bright.

2.5 T2* and the spin-echo pulse cycle

In the previous sections, it was implied that the decay of the MDM.s in the x-y plane after termination of the 90^0 RF-pulse equals the T2 relaxation signal. This is, however, a simplification and to really understand the use of MRI we will need to explain T2 in more detail. True T2 decay is actually a lot slower than the decay of the MDM.s in the x-y plane after termination of the 90^0 RF-pulse. The decay of the MDM.s in the x-y plane is more accurately

described as T2* decay [8] [9]. The reason why the MDM.s decay in the x-y plane (the T2* signal) is essentially due to dephasing [8] [9]. Remember that the MDM.s of hydrogen nuclei in a magnetic field of a certain field strength all precess at the same frequency. However, before the application of the 90^0 RF-pulse, they are not precessing in the same phase.

When applied to the precession of the MDM.s in a magnetic field, this means that even though the MDM.s precess at the same frequency, they will each be at a different position in their cycle at a given point in time i.e. they are precessing in a different phase. As shown in Figure 2.3 the top row displays one moment in the cycle of precession of 3 MDM.s that are in phase & the bottom row displays one moment in the cycle of precession of 3 MDM.s that are not in phase.

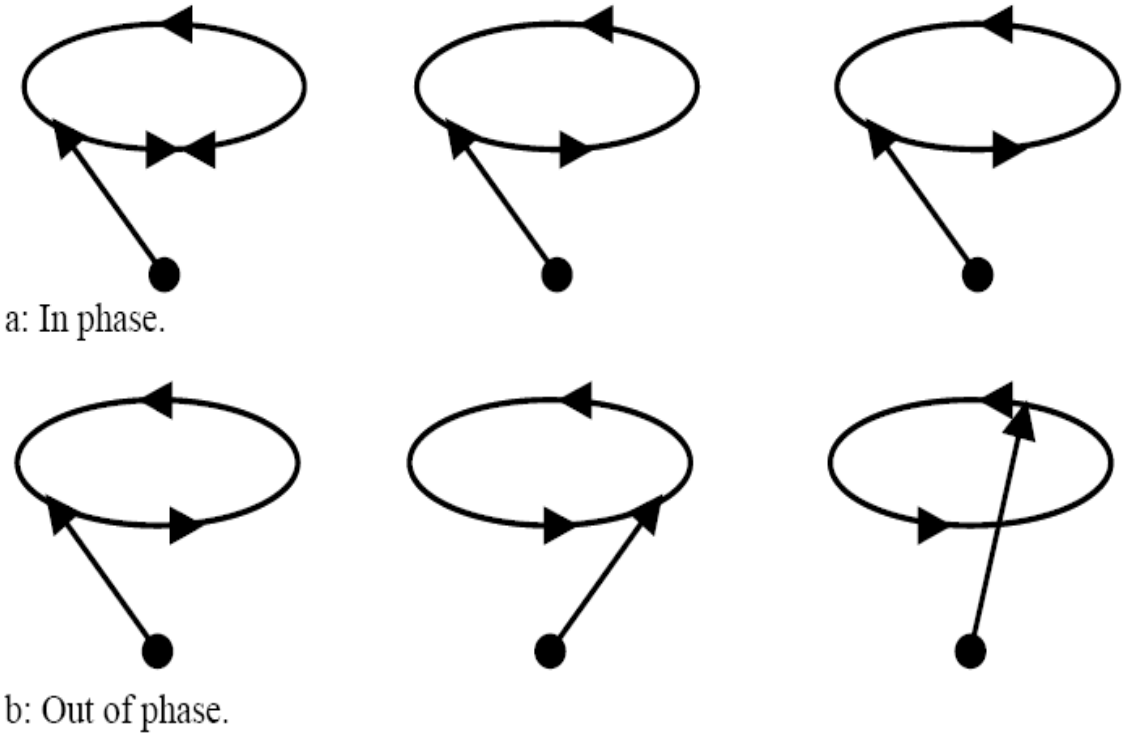


Figure 2.3: Cycle of precession of 3 MDM.s that are in phase & not in phase

At the moment when the 90^0 RF-pulse is applied, the MDM.s are tipped in the x-y plane and are forced to precess in phase (as the brief RF-pulse tips all the MDM.s simultaneously). Now the MDM signals are additive and therefore result in a strong signal in the x-y plane.

After termination of the 90^0 RF-pulse, however, the precession of the MDM.s will gradually dephase. The MDM signals are now no longer additive, but cancel each other out and the signal decays. This is the T2* decay [8] [9].

There are two reasons why the MDM.s dephase after termination of the 90^0 RF-pulse and the signal decays [8] [9]. The first one is magnetic field inhomogeneity. The strength of the magnetic field is not uniform and since the frequency of precession of the MDM.s depends on the strength of the magnetic field it follows that different MDM.s will precess at a different frequency and therefore this precession will get out of phase. The second reason why the MDM.s dephase is because of spin-spin interaction [8] [9].

Different hydrogen nuclei are surrounded by different other nuclei. These other nuclei affect the frequency of precession of the MDM of the hydrogen nuclei. The frequency of precession of each MDM.s will be differently affected by the surrounding nuclei. This, again, results in different frequencies of precession for different MDM.s and hence dephasing occurs [8] [9].

To summarize, initially the precession of the MDM.s is dephased. When the 90^0 RFpulse is applied, the MDM.s are forced to precess in phase, resulting in a signal in the x-y plane. After termination of the 90^0 RF-pulse, the precession of the MDM.s will dephase again due to the inhomogeneities in the magnetic field and spin-spin interactions and the signal in the x-y plane decays. This is T2* decay.

The trick is that the dephasing due to the inhomogeneity of the magnetic field is correctable and by correcting for this source of dephasing the true T2 signal is obtained [8] [9]..Directly after application of the 90^0 RF-pulse the precession of the MDM.s is in phase.

This means that at a given time, all the MDM.s will be at the same point in their cycle of precession. After termination of the 90^0 RF-pulse the precession of the MDM.s will slowly

dephase. Basically, at a given time, different MDM.s will not longer be at the same point in their cycle of precession.

If an RF-pulse is now applied from the opposite direction (180^0) as the direction from which the original RF-pulse was applied (90^0), the direction of rotation of the precession of the MDM.s is reversed.

After the same amount of time has elapsed following the 180^0 RF-pulse as the amount of time between the 90^0 RF-pulse and the 180^0 RF-pulse, the MDM.s will be in phase again [8] [9]. After application of the 90^0 RF-pulse all the MDM.s are at the same point in their cycle of precession. After termination of the RF-pulse some MDM.s will rotate faster (have higher frequencies of precession) than others, the MDM.s dephase.

After a while the 180^0 RF-pulse is applied and this makes the direction of precession of all the MDM.s reverse. The MDM.s, however, all keep their own frequency of precession and will arrive at their starting point (the point where they were directly after application of the 90^0 RF-pulse) in the cycle of precession at the same time, they will be in phase again.

The important thing is that the time between the 180^0 RF-pulse and measurement of the signal must be the same as the time between the 90^0 RF-pulse and the 180^0 RF-pulse. A measurement cycle where one 90^0 RF-pulse is followed by one or more 180^0 RF-pulses with a measurement after each 180^0 RF-pulse is known as a spin-echo pulse cycle.

Even though the spin-echo pulse cycle corrects for the decay in the signal caused by the inhomogeneities in the magnetic field, the signal still does eventually decay because of the dephasing due to spinspin interactions. The decay is now, however, a lot slower and this is the true T2 decay [8] [9].

2.6 Overview of MRI Imaging sessions

There are three primary steps to an MRI Imaging procedure: preparation, acquisition and processing.

2.6.1 Preparation:

The patient is given a detailed explanation of the procedure and carefully instructed on the chosen task in order to obtain the highest quality exam with the least amount of patient-induced motion. It is common to practice these tasks with the patient before the exam to assure the best results.

2.6.2 Acquisition:

The first step in the acquisition process is to collect routine 3D datasets. These will later be used as the anatomical data upon which the MRI information will be mapped. Activity in specific regions of the brain is induced and controlled by a set of tasks called a paradigm. These tasks are performed by the patient during the BOLD imaging measurements. MRI data collection is done in three sessions described below.

The following table illustrates the overview of imaging sessions for MRI data collection

Session 1: Structural & functional sequences	**Duration**
1.Volunteer preparation / equipment adjustment	20:00
2. 3 plane localizer / Parallel imaging calibration	00:22
3. Axial T2 slices (site specific duration)	~01:19
4. Axial T2 Flair slices (site specific duration)	~ 02:25
5. Instructions / talk to volunteer	2:00
6. Face task	5:00
7. Instructions / talk to volunteer	2:00
8. Stop-signal task	16:00
9. B0 Map	00:40
10. 3D Sagittal ADNI MPRAGE (Long)	09:17
Duration	**~60 min**

Session 2: Structural & Functional sequences	**Duration**
1.Volunteer preparation / equipment adjustment	14:00
2.3 plane localizer / Parallel imaging calibration	00:22
3. B0 Fieldmap	00:40
4. 3D Sagittal ADNI MPRAGE (Short)	02:23
5. Instructions / talk to volunteer	2:00
6. M&M Incentive Delay Task	11:06
7. Instructions / talk to volunteer	02:00
8. Global Cognition Assessment	05:00
9. Instructions / talk to volunteer	02:00
10. Breath Hold Task (optional)	05:40
11. DTI (duration is heart-rate dependent at sites with cardiac gating)	10:00
Duration	**~60 min**

Optional Session 3: Structural & functional sequences	**Duration**
1.3 plane localizer / Parallel imaging calibration	00:22
2. Despot	18:30
3. 3D Sagittal ADNI MPRAGE (Short)	02:23
Duration	**~22 min**

Table 2.1: Overview of imaging sessions for MRI data collection

The various tasks performed during the coarse of imaging are

a. Face task

In this task volunteers are asked to passively watch video clips presenting faces with neural and angry expressions as well as control non-biological motion stimuli (concentric circles). After scanning a short recognition test is performed outside the scanner with 5 static pictures extracted from the movies.

b. Stop-signal Task

The main principle of this task is to respond to regular presented visual go stimuli (go trials) but to withhold the motor response to the go stimulus when it is followed unpredictably by a stop signal (stop trials). This task yields an estimate of a subject's stop-signal reaction time (SSRT). The SSRT is thought to be directly reflective of the central inhibitory mechanism.

c. MID (M&M Incentive Delay) Task

This task is a reaction time task - it tests how quickly the subject can react and pull the trigger to hit a target (with left or right index finger) that only appears for a short time on the left or right of the screen. If the subject can hit the target, they will score points. The subject can tell where the target will appear and how many points they can win by the symbol they see on the screen before each trial. A triangle means no points, a circle with a line means 2 points and a circle with three lines means 10 points. Responding too early or too late will result in a loss.

d. Global Cognition Assessment Task

This task is composed by the following brief tasks:
1. Passive viewing of a flashing checkerboard (20 trials)
2. Pressing three times the left button with the left index finger according to visual instructions (5 trials)
3. Pressing three times the right button with the right index finger according to visual instruction (5 trials)

4. Pressing three times the left button with the left index finger according to auditory instruction (5 trials)

5. Pressing three times the right button with the right index finger according to auditory instruction (5 trials)

6. Reading silently short visual sentences (10 trials)

7. Listening to short sentences (10 trials)

8. Solving silently visual subtraction problems (10 trials)

9. Solving silently auditory subtraction problems (10 trials)

e. Breath Hold Task (Paced Expiration Breath Hold Task)

This task uses visual instructions to pace their breathing in a regular rhythm for 40 seconds (breathing in for 4 seconds and out for 4 seconds), followed by holding their breath on expiration for a short periods of 20 seconds. This cycle is then repeated five times, ending on paced breathing to give a total task length of 5 minutes 40 seconds. This task uses the small build up of carbon dioxide to assess vascular responsivity in each participant and which differs between participants.

2.6.3 Processing:

After data collection, a statistical evaluation (t-test) is used to generate BOLD maps that are combined with routine 3D imaging datasets such as MPRAGE. The combined data can then be used as a neuronavigational roadmap for use in pre surgical planning or treatment assessment.

CHAPTER 3

MRI IMAGE PROCESSING

This Chapter introduces the various methods used for the processing of MRI image. The MRI image analysis is performed under following sequence of operations:

1. Features Extraction Using Wavelet Transform
2. Features Reduction Using Principal Component Analysis
3. Classification using Artificial Neural Network

The detailed explanation of all above operations is presented here.

3.1 Wavelet Transform

The transform of a signal is just another form of representing the signal. It does not change the information content present in the signal. The Wavelet Transform provides a time-frequency representation of the signal. It was developed to overcome the short coming of the Short Time Fourier Transform (STFT), which can also be used to analyze non-stationary signals. While STFT gives a constant resolution at all frequencies, the Wavelet Transform uses multi-resolution technique by which different frequencies are analyzed with different resolutions. A wave is an oscillating function of time or space and is periodic. In contrast, wavelets are localized waves. They have their energy concentrated in time or space and are suited to analysis of transient signals. While Fourier Transform and STFT use waves to analyze signals, the Wavelet Transform uses wavelets of finite energy.

The wavelet analysis is done similar to the STFT analysis. The signal to be analyzed is multiplied with a wavelet function just as it is multiplied with a window function in STFT, and then the transform is computed for each segment generated. However, unlike STFT, in Wavelet Transform, the width of the wavelet function changes with each spectral component The Wavelet Transform, at high frequencies, gives good time resolution and poor frequency resolution, while at low frequencies, the Wavelet Transform gives good frequency resolution and poor time resolution.

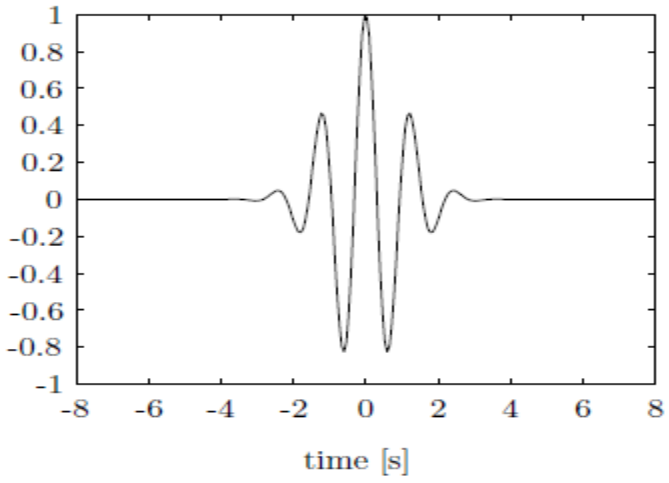


Figure 3.1 Morlet Wavelet

An analyzing function Ψ(t) is classified as a wavelet if the following mathematical criteria are satisfied:

1. A wavelet must have finite energy

$$E = \int_{-\infty}^{\infty} \left|\Psi(t)\right|^2 dt \langle \infty \tag{3.1}$$

The energy E equals the integrated squared magnitude of the analyzing function Ψ(t) and must be less than infinity.

2. If Ψ (f) is the Fourier transform of the wavelet Ψ (t), the following condition must hold

$$C_\psi = \int_0^{\infty} \frac{\left|\hat{\Psi}(f)\right|^2}{f} df \langle \infty \tag{3.2}$$

This condition implies that the wavelet has no zero frequency component (Ψ (0) = 0), i.e. the mean of the wavelet Ψ (t) must equal zero. This condition is known as the admissibility constant. The value of C_Ψ depends on the chosen wavelet.

3. For complex wavelets the Fourier transform Ψ (f) must be both real and vanish for negative frequencies.

3.1.1 Continuous Wavelet Transform

We define a mother wavelet function $\Psi(t) \in L^2(R)$, which is limited in time domain. That is, $\Psi(t)$ has values in a certain range and zeros elsewhere. Another property of mother wavelet is zero mean. The other property is that the mother wavelet is normalized. Original lectures can be found in [10]. Mathematically, they are

$$\int_{-\infty}^{\infty} \Psi(t)dt = 0 \tag{3.3}$$

$$\left\|\Psi(t)\right\|^2 = \int_{-\infty}^{\infty} \Psi(t)\Psi^* dt = 1 \tag{3.4}$$

As the dilation and translation property states, the mother wavelet can form a basis set denoted by

$$\left\{\Psi s,u(t) = \frac{1}{\sqrt{s}}\Psi\left(\frac{t-u}{s}\right)\right\}\Big|_{u\in R, s\in R+} \tag{3.5}$$

u is the translating parameter, indicating which region we concern. s is the scaling parameter greater than zero because negative scaling is undefined. The multiresolution property ensures the obtained set $\{\Psi_{u,s}(t)\}$ is orthonormal. Conceptually, the continuous wavelet transform is the coefficient of the basis $\Psi_{u,s}(t)$ [11] . It is

$$Wf(s,u) = \left\langle f(t), \Psi_{s,u} \right\rangle \tag{3.6}$$

$$= \int_{-\infty}^{\infty} f(t)\Psi_{s,u}^*(t)dt \tag{3.7}$$

$$= \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{s}}\Psi^*(\frac{t-u}{s})dt \tag{3.8}$$

The Continuous Wavelet Transform (CWT) is provided by equation 3.8, where f(t) is the signal to be analyzed. $\Psi(t)$ is the mother wavelet or the basis function. All the wavelet functions used in the transformation are derived from the mother wavelet through translation (shifting) and scaling (dilation or compression). The transformed signal W f(s,u) is a function of the translation

parameter τ and the scale parameter s. The mother wavelet is denoted by Ψ, the * indicates that the complex conjugate is used in case of a complex wavelet. The signal energy is normalized at every scale by dividing the wavelet coefficients by $1/\sqrt{|s|}$. This ensures that the wavelets have the same energy at every scale. The mother wavelet is contracted and dilated by changing the scale parameter s. The variation in scale s changes not only the central frequency fc of the wavelet, but also the window length. Therefore the scale s is used instead of the frequency for representing the results of the wavelet analysis. The translation parameter u specifies the location of the wavelet in time, by changing u the wavelet can be shifted over the signal. For constant scale s and varying translation u the rows of the time-scale plane are filled, varying the scale s and keeping the translation u constant fills the columns of the time-scale plane. The elements in Wf(s,u) are called wavelet coefficients, each wavelet coefficient is associated to a scale (frequency) and a point in the time domain.

Via this transform, one can map an one-dimensional signal f(t) to a two-dimensional coefficients Wf(s,u). The two variables can perform the time frequency analysis. We can tell locate a particular frequency (parameter s) at a certain time instant (parameter u). If the f(t) is a $L^2(R)$ function. The inverse wavelet transform is

$$f(t) = \frac{1}{C_\Psi} \int_0^\infty \int_{-\infty}^\infty Wf(s,u) \frac{1}{\sqrt{s}} \Psi\left(\frac{t-u}{s}\right) du \frac{ds}{s^2} \tag{3.9}$$

where C_Ψ is defined as

$$C_\Psi = \int_0^\infty \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty \tag{3.10}$$

Ψ(ω) is the Fourier transform of the mother wavelet Ψ(t). This equation is also called the admissibility condition Note that the admissibility constant C must satisfy the second wavelet condition. A wavelet function has its own central frequency f_c at each scale, the scale s is inversely proportional to that frequency. A large scale corresponds to a low frequency, giving global information of the signal. Small scales correspond to high frequencies, providing detail signal information. The mother wavelet used to generate all the basis functions is designed based on some desired characteristics associated with that function. The translation parameter τ relates

to the location of the wavelet function as it is shifted through the signal. Thus, it corresponds to the time information in the Wavelet Transform. The scale parameter *s* is defined as |1/frequency| and corresponds to frequency information.

Scaling either dilates (expands) or compresses a signal. Large scales (low frequencies) dilate the signal and provide detailed information hidden in the signal, while small scales (high frequencies) compress the signal and provide global information about the signal. Notice that the Wavelet Transform merely performs the convolution operation of the signal and the basis function. The above analysis becomes very useful as in most practical applications, high frequencies (low scales) do not last for a long duration, but instead, appear as short bursts, while low frequencies (high scales) usually last for entire duration of the signal.

The Wavelet Series is obtained by discretizing CWT. This aids in computation of CWT using computers and is obtained by sampling the time-scale plane. The sampling rate can be changed accordingly with scale change without violating the Nyquist criterion. Nyquist criterion states that, the minimum sampling rate that allows reconstruction of the original signal is 2ω radians, where ω is the highest frequency in the signal. Therefore, as the scale goes higher (lower frequencies), the sampling rate can be decreased thus reducing the number of computations.

3.1.2 Discrete Wavelet Transform:

The Wavelet Series is just a sampled version of CWT and its computation may consume significant amount of time and resources, depending on the resolution required. The Discrete Wavelet Transform (DWT), which is based on sub-band coding is found to yield a fast computation of Wavelet Transform. It is easy to implement and reduces the computation time and resources required. The foundations of DWT go back to 1976 when techniques to decompose discrete time signals were devised. Similar work was done in speech signal coding which was named as sub-band coding. In 1983, a technique similar to sub-band coding was developed which was named pyramidal coding. Later many improvements were made to these coding schemes which resulted in efficient multi-resolution analysis schemes.

In CWT, the signals are analyzed using a set of basis functions which relate to each other by simple scaling and translation. In the case of DWT, a time-scale representation of the digital

signal is obtained using digital filtering techniques. The signal to be analyzed is passed through filters with different cutoff frequencies at different scales.

3.1.2.1 Filter banks

A filter bank consists of filters which separate a signal into frequency bands [12]. An example of a two channel filter bank is shown in Fig.3.2. A discrete time signal x(k) enters the analysis bank and is filtered by the filters L(z) and H(z) which separate the frequency content of the input signal in frequency bands of equal width. The filters L(z) and H(z) are therefore respectively a low-pass and a high-pass filter. The output of the filters each contains half the frequency content, but an equal amount of samples as the input signal. The two outputs together contain the same frequency content as the input signal, however the amount of data is doubled. Therefore down sampling by a factor two, denoted by ↓ 2, is applied to the outputs of the filters in the analysis bank. Reconstruction of the original signal is possible using the synthesis filter bank [12] [13]. In the synthesis bank the signals are up sampled (↑2) and passed through the filters L′(z) and H′(z). The filters in the synthesis bank are based on the filters in the analysis bank. The outputs of the filters in the synthesis bank are summed, leading to the reconstructed signal y(k). The different output signals of the analysis filter bank are called subbands, the filter-bank technique is also called subband coding [13].

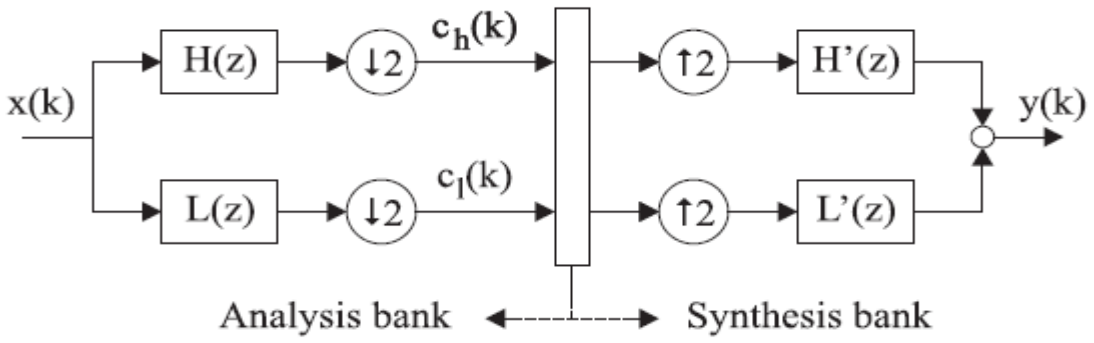


Figure 3.2: Two channel filter bank

3.1.2.2 Conditions for Perfect Reconstruction

For perfect reconstruction to be possible, the filter bank should be biorthogonal. Furthermore some design criteria for both the analysis and synthesis filters should be met to prevent aliasing and distortion and to guarantee a perfect reconstruction [12].

In the two channel filter bank of Fig. 3.2, the filters L(z) and H(z) split the signal into two frequency bands, i.e. the filters are respectively a low-pass and a high-pass filter. If the filters were perfect brick-wall filters, the down sampling would not lead to loss of information. However ideal filters cannot be realized in practice, so a transition band exists. Besides aliasing, this leads to an amplitude and phase distortion in each of the channels of the filter band [13]. For the two channel filter bank of Fig. 3.2, aliasing can be prevented by designing the filters of the synthesis filter bank as [12]

$$L'(z) = H(-z) \tag{3.11}$$

$$H'(z) = -L(-z) \tag{3.12}$$

To eliminate distortion, a product filter $P_0(z) = L'(z)L(z)$ is defined. Distortion can be avoided if

$$P_0(z) - P_0(-z) = 2z^{-N} \tag{3.13}$$

where N is the overall delay in the filter bank. Generally an N^{th} order filter produces a delay of N samples [13]. The perfect reconstruction filter bank can be designed in two steps [12]:

1. Design a low-pass filter P_0 satisfying (3.13).
2. Factor P_0 into $L'(z)L(z)$ and use (3.11) and (3.12) to calculate $H(z)$ and $H'(z)$.

3.1.2.3 Multi-Resolution Analysis using DWT Filter Banks

The CWT of Chapter 3 performs a multiresolution analysis which makes it possible to analyze a signal at different frequencies with different resolutions. For high frequencies (low scales), which last a short period of time, a good time resolution is desired. For low frequencies (high scales) a good frequency resolution is more important. The CWT has a time-frequency resolution. This multiresolution can also be obtained using filter banks, resulting in the discrete wavelet transform (DWT). Note that the discretized version of the CWT is not equal to the DWT, the DWT uses filter banks, whereas the discretized CWT uses discretized versions of the scale and dilatation axes.

The low-pass and high-pass filtering branches of the filter bank retrieve respectively the approximations and details of the signal x(k). In Fig. 3.3, a three level filter bank is shown. The filter bank can be expanded to an arbitrary level, depending on the desired resolution. The coefficients $c_l(k)$ represent the lowest half of the frequencies in x(k), downsampling doubles the frequency resolution. The time resolution is halved, i.e. only half the number of samples are present in $c_l(k)$. In the second level, the outputs of L(z) and H(z) double the time resolution and decrease the frequency content, i.e. the width of the window is increased. After each level, the output of the high-pass filter represents the highest half of the frequency content of the low-pass filter of the previous level, this leads to a pass-band. For a special set of filters L(z) and H(z) this structure is called the DWT, the filters are called wavelet filters.

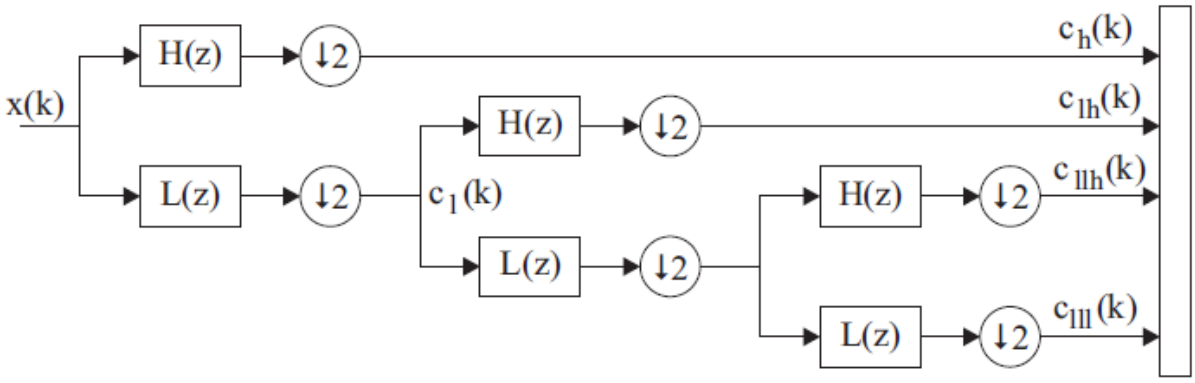


Figure 3.3: Three level analysis filter bank

At each decomposition level, the half band filters produce signals spanning only half the frequency band. This doubles the frequency resolution as the uncertainty in frequency is reduced by half. In accordance with Nyquist's rule if the original signal has a highest frequency of ω, which requires a sampling frequency of 2ω radians, then it now has a highest frequency of ω/2 radians. It can now be sampled at a frequency of ω radians thus discarding half the samples with no loss of information. This decimation by 2 halves the time resolution as the entire signal is now represented by only half the number of samples. Thus, while the half band low pass filtering removes half of the frequencies and thus halves the resolution, the decimation by 2 doubles the scale. With this approach, the time resolution becomes arbitrarily good at high frequencies, while the frequency resolution becomes arbitrarily good at low frequencies. The filtering and decimation process is continued until the desired level is reached. The maximum number of

levels depends on the length of the signal. The DWT of the original signal is then obtained by concatenating all the coefficients, a[n] and d[n], starting from the last level of decomposition.

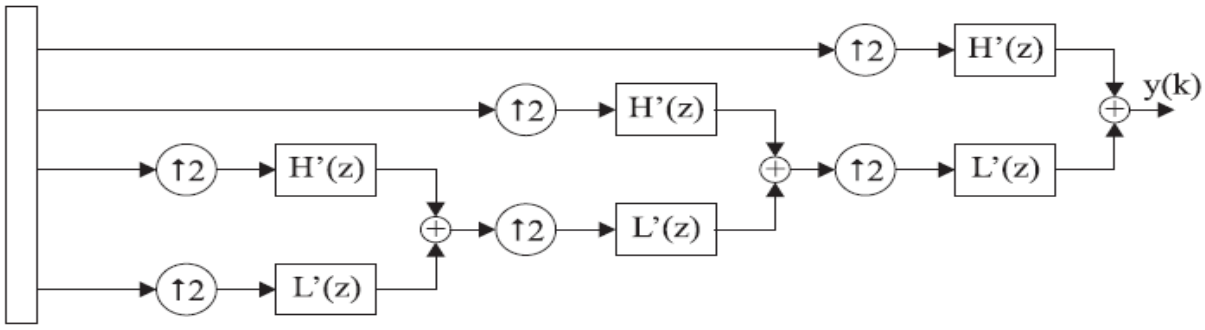


Figure 3.4: Three level synthesis filter bank

Figure 3.4 shows the reconstruction of the original signal from the wavelet coefficients. Basically, the reconstruction is the reverse process of decomposition. The approximation and detail coefficients at every level are up sampled by two, passed through the low pass and high pass synthesis filters and then added. This process is continued through the same number of levels as in the decomposition process to obtain the original signal. Mallat [14] shows that wavelet transform is implemented by using two quadrature mirror filters H (low-pass) and G (high-pass) The Mallat algorithm works equally well if the analysis filters, $G_0$and H_0 , are exchanged with the synthesis filters, G_1 and H_1.

3.1.3 Segmentation & Features Extraction Using Discrete Wavelet transform in 2-D Image

After pre-processing phase, we adopt a segmentation algorithm. The basic aim of segmentation is the separation of an image into homogeneous regions (spatially connected groups of pixels called classes, or subsets) with respect to one or more characteristics or features; such that the union of any two neighboring regions yields a heterogeneous.

Medical image segmentation is a promising field and imposes constraints related to the concept of time, the great number of implied data and the richness of image concerning the complexity of the organ's anatomy, the patient's position of catching image. All these medical images characteristics add more difficulties to the problem of image segmentation and make the construction of a general model more complex. This explains the variety of segmentation

methods appeared in the last years. In literature there exist two major classes of segmentation techniques: edge based segmentation approach and region based segmentation approach.

Edge approach looks for limits between regions with different characteristics. Its aims at finding object boundaries and segmenting regions enclosed by the contours. There are various Edge based techniques widely used for edge based segmentation approach include Roberts, Prewitt, Robinson, Kirsch, and Laplacian [15]. They prove to be computationally fast and don't require prior information about the image content. However a drawback of the edge approach is that the edges do not enclose the object completely. In region-based techniques, segmentation is applied by identifying all pixels that belong to the object based on the intensity of pixels. They include region growing, watershed algorithm and thresholding [16]. More recently, with the application of a spatial-frequency image analysis, multiscale techniques have sparked the interest of researchers for segmentation of images especially wavelet transform.

Wavelet transform represents an image at different resolution levels. At resolution j, it provides an approximation of the original image I_j and three detail of image $\mathbf{D_2^V}$, $\mathbf{D_2^H}$, $\mathbf{D_2^D}$ .Each of these details images privileges a particular orientation: horizontal, vertical and diagonal, and preserves the lost information during their passage from j-1 to j. Figure 3.5 illustrated a decomposition of image at two levels. It has been shown that the wavelet coefficients resulting from this transformation contain the information concerning the original image for different scales [17]. Wavelet coefficients represent the degree of correlation or similarity between the image and the mother wavelet at the particular scale and translation. Thus the set of all wavelet coefficients gives the wavelet domain representation of the image.

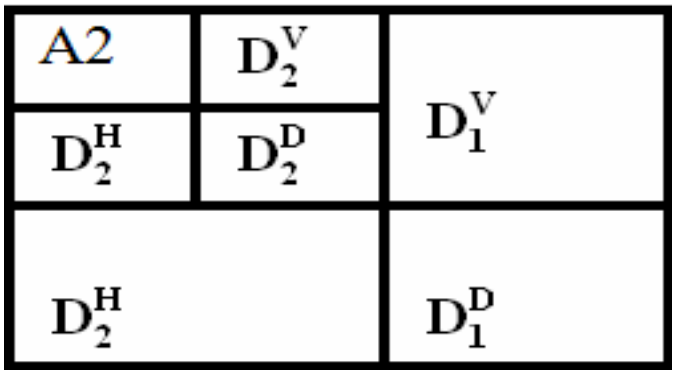


Figure 3.5: Sub-images generated at two levels

Wavelet coefficients are used for generating the initial features. Wavelet transform is traditionally used for feature extraction. The provision of localized frequency information about a function of a signal is the main advantage of wavelets and is particularly beneficial for classification. Wavelets have been used as a feature extraction method for discrimination.

In two-dimensional wavelet transform a scaling function $\Phi(x,y)$and three wavelets $\Psi^H(x,y)$ measures variations along columns), $\Psi^V(x,y)$ (responds to variations along rows), $\Psi^D(x,y)$ (corresponds to variations along diagonals), are required. The discrete wavelet transform of image f(x,y) of size MXN is then

$$W_\varphi(j_0,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\varphi_{j_0,m,n}(x,y) \tag{3.14}$$

$$W_\psi^i(j,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\varphi_{j,m,n}^i(x,y) \tag{3.15}$$

Where i identifies the directional wavelets (i= {H, V, D}) and J_0 is an arbitrary starting scale. The $W_\Phi(J_0,m,n)$ coefficients define an approximation of f(x,y) and scale j_0 .The $W^i\Psi(j,m,n)$ coefficients add horizontal, vertical, and diagonal details for scales $j>=j_0$[18]. Fig.2.6 shows the process in block diagram form. There are several different kinds of wavelets which have gained popularity throughout the development of wavelet analysis. One important discrete wavelet is the Haar wavelet. Basically, it is one period of a square wave. Because of its simplicity, it is often the wavelet to be chosen [19]. Fig.3.6 shows the discrete wavelet transform of one example MR image.

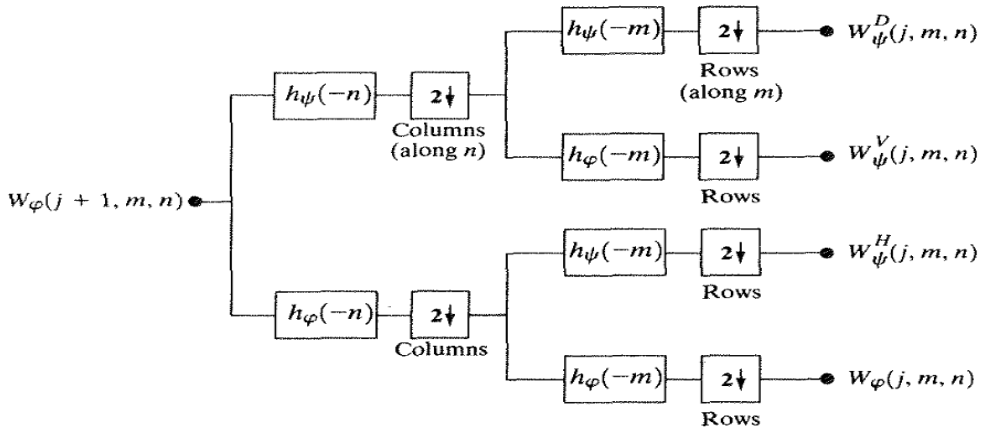


Figure 3.6: The analysis filter banks of discrete wavelet transform

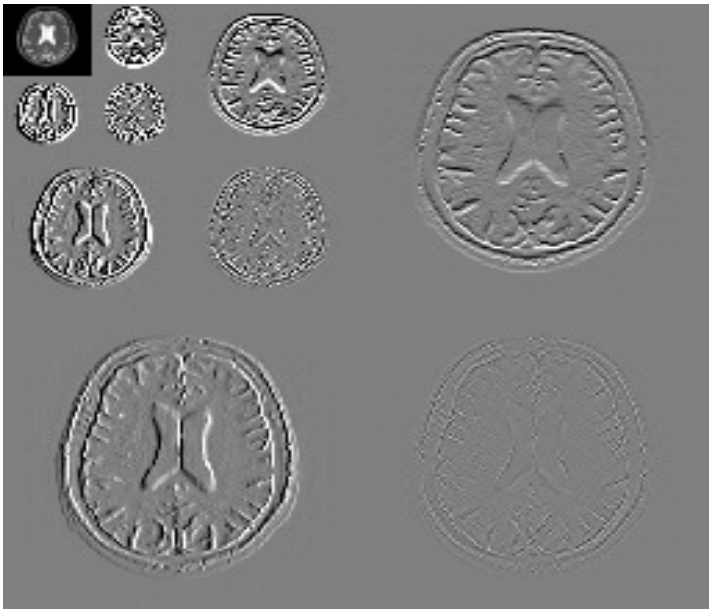


Figure 3.7: Discrete wavelet transform of one example MR image

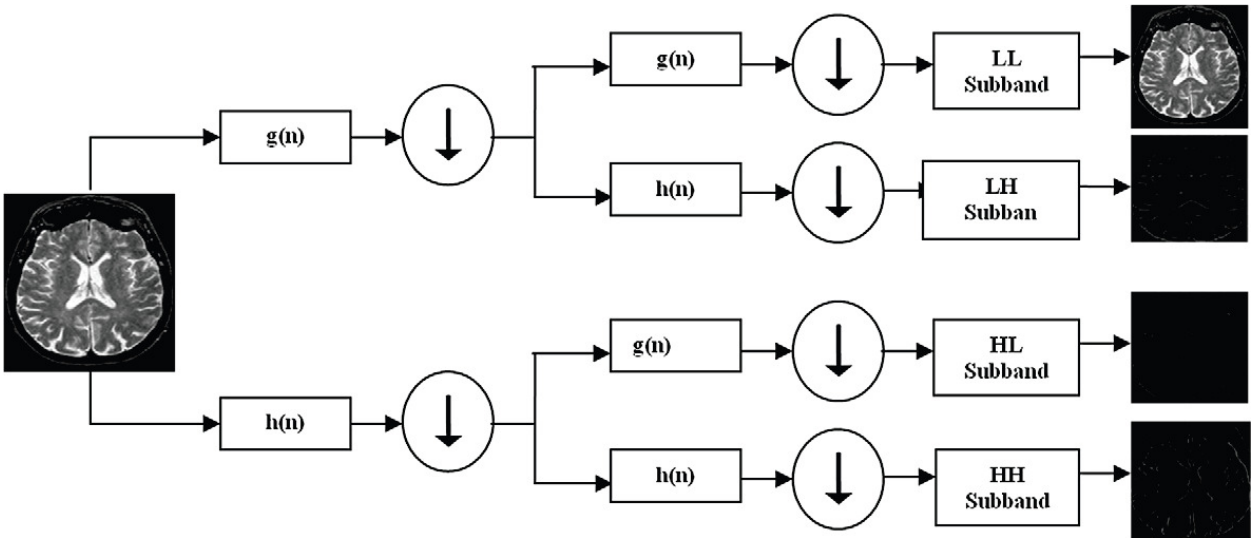


Figure 3.8: DWT schematically [20]

3.1.4 Wavelet Families

There are a number of basis functions that can be used as the mother wavelet for Wavelet Transformation. Since the mother wavelet produces all wavelet functions used in the transformation through translation and scaling, it determines the characteristics of the resulting Wavelet Transform. Therefore, the details of the particular application should be taken into account and the appropriate mother wavelet should be chosen in order to use the Wavelet Transform effectively.

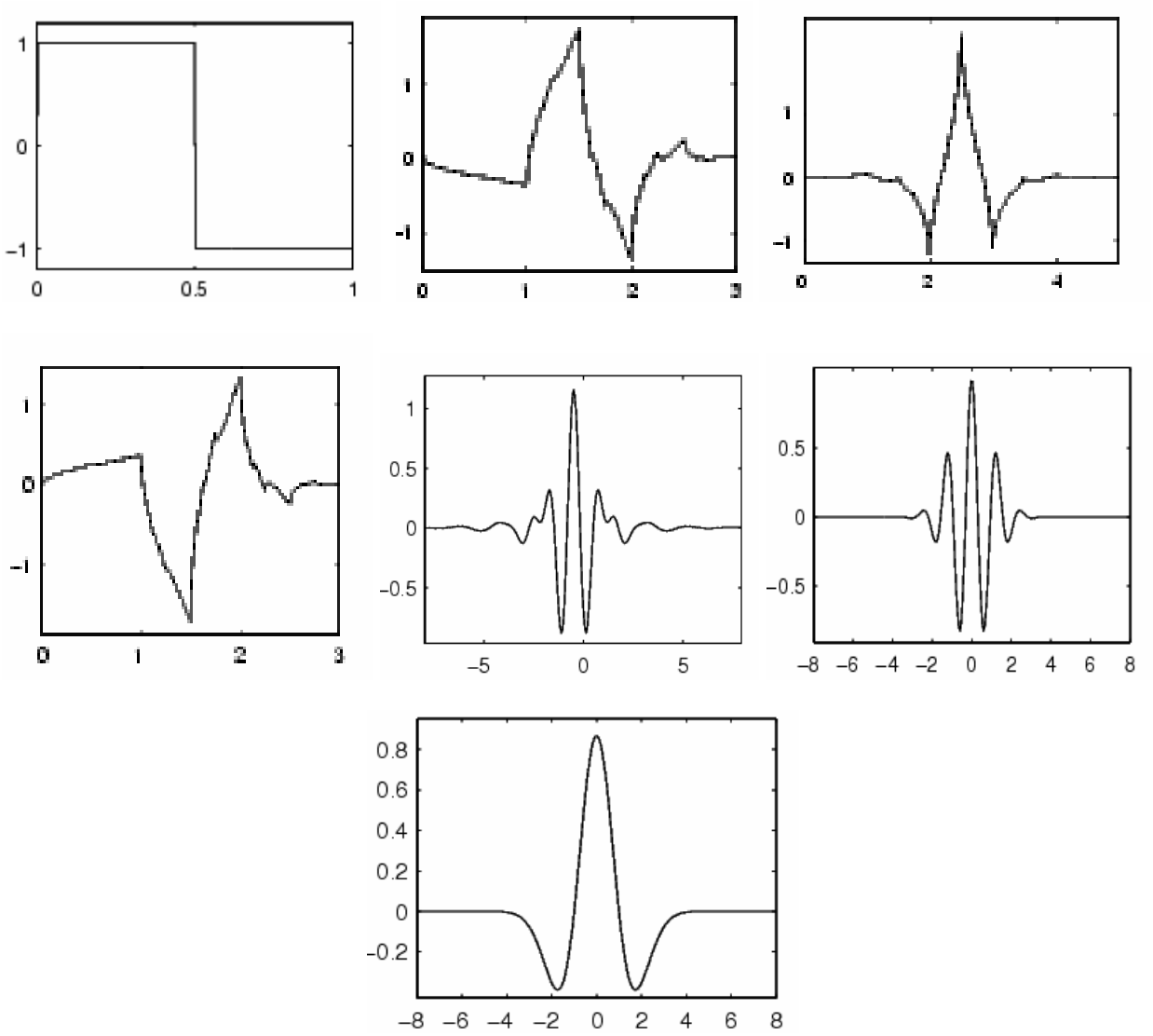


Figure 3.9: Wavelet families (a) Haar (b) Daubechies4 (c) Coiflet1 (d) Symlet2 (e) Meyer (f) Morlet (g) Mexican Hat.

3.2 Principal Component Analysis

Measurement cost and classification accuracy are two predominant reasons for minimizing the dimensionality of the pattern representation (i.e., the number of features). Classifiers which are built on the selected representation can use less memory and be faster by utilizing the limited feature set. Linear transforms, due to their simplicity, have been widely used for feature extraction. These transforms create a smaller set of features from linear combination of the initial features. One of the best known linear feature extractor is the principal component analysis (PCA). PCA is a technique for simplifying a data set, by reducing multidimensional data sets to lower dimensions for analysis.

PCA is an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA can be used for dimension reduction in a data set while retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the "most important" information of the data. But this is not necessarily the case, depending on the application.

PCA involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called "principal components". The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Principal component analysis (Karhunen-Loeve or Hotelling transform) [21] - PCA belongs to linear transforms based on the statistical techniques. This method provides a powerful tool for data analysis and pattern recognition which is often used in signal and image processing [18] as a technique for data compression, data dimension reduction or their decorrelation as well.

There are various algorithms based on multivariate analysis or neural networks that can perform PCA on a given data set. Excessive features increase computation time and storage memory which sometimes causes the classification process to become more complicated. This

consequence is called the curse of dimensionality. A strategy is necessary to reduce the number of features used in classification.

PCA is an efficient tool to reduce the dimension of a data set consisting of a large number of interrelated variables while retaining the most significiant variations. It is achieved by transforming the data set to a new set of ordered variables according to their degree of variance or importance. This technique has three effects:

(i) It orthogonalizes the components of the input vectors so that they are uncorrelated with each other.
(ii) It orders the resulting orthogonal components so that those with the largest variation come first.
(iii) It eliminates the components in the data set that contributing the least variation.

3.2.1 PCA Theory

Principal component analysis in signal processing can be described as a transform of a given set of n input vectors (variables) with the same length K formed in the n-dimensional vector $x = [x_1, x_2, ...x_n]^T$ into a vector y according to

$$y = A\,(x - m_x) \tag{3.16}$$

This point of view enables to form a simple formula (3.16) but it is necessary to keep in the mind that each row of the vector **x** consists of K values belonging to one input. The vector $\mathbf{m_X}$ in Eq. (3.17) is the vector of mean values of all input variables defined by relation

$$m_X = E\{X\} = \frac{1}{K}\sum_{k=1}^{K} X_k \tag{3.17}$$

Matrix **A** in Eq. (3.16) is determined by the covariance matrix **Cx**. Rows in the **A** matrix are formed from the eigenvectors **e** of **Cx** ordered according to corresponding eigenvalues in descending order. The evaluation of the **Cx** matrix is possible according to relation

$$C_X = E\{(X - m_X)(X - m_X)^T\} = \frac{1}{K}\sum_{k=1}^{K} X_k X_k^T - m_X m_X^T \tag{3.18}$$

As the vector **x** of input variables is n-dimensional it is obvious that the size of $\mathbf{C_X}$ is n x n. The elements **Cx**(i, i) lying in its main diagonal are the variances

$$C_X(i,i) = E\{ (x_i - m_i)^2 \} \tag{3.19}$$

of **x** and the other values **Cx**(i, j) determine the covariance between input variables $\mathbf{x_i}$, $\mathbf{x_j}$.

$$C_X(i,j) = E\{(x_i - m_i)(x_j - m_j)\} \tag{3.20}$$

between input variables x_i, x_j . The rows of **A** in Eq. (3.16) are orthonormal so the inversion of PCA is possible according to relation

$$x = A^T y + m_x \tag{3.21}$$

The kernel of PCA defined by Eq. (3.21) has some other interesting properties resulting from the matrix theory which can be used in the signal and image processing [22] .

In PCA, the input feature space is transformed into a lower-dimensional feature space using the largest eigenvectors of the correlation matrix. PCA is the most widely used subspace projection technique. When a set of data is given, PCA finds the linear lower-dimensional representation of the data such that the variance of the data is preserved. Using a system of feature reduction based on PCA limits the feature vectors to the component selected by the PCA which leads to an efficient classification algorithm. So, the main idea behind using PCA is to reduce the dimensionality of the wavelet coefficients which results in a more efficient and accurate classifier.

These methods provide suboptimal solution with a low computational cost and computational complexity. Given a set of data, PCA finds the linear lower-dimensional representation of the data such that the variance of the reconstructed data is preserved. Using a system of feature reduction based on PCA limits the feature vectors to the component selected by the PCA which leads to an efficient approach to reduce the dimensionality of the wavelet coefficients which results in a more efficient and accurate classifier.

PCA "squeezes" as much information (as measured by variance) as possible into the first principal components. In some cases the number of principal components needed to store the vast majority of variance is shockingly small: a tremendous feat of data manipulation. This transformation can be performed quickly on contemporary hardware and is invertible, permitting any number of useful applications. The following algorithm is used to find out the principal components of the input matrix to the classifier. Now the input matrix consists of only these principal components.

3.2.2 PCA algorithm

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

Perform the following steps [20]:

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

2. Calculate the deviations from the mean and store the data in the matrix

B[M × N]: B = X -u · h, where h is a 1 × N row vector of all 1's: h[n] = 1 for n = 1, . . . , N.

3. Find the covariance matrix C: C = B · B*

4. Find the eigenvectors and eigenvalues of the covariance matrix $V^{-1}CV = D$

V – the eigenvectors matrix;

D – the diagonal matrix of eigenvalues of C,

D[p, q] = λ_m for p=q=m is mth eigen value of covariance matrix C.

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

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, . . . ,M, q = 1, . . . , L where $1 \leq L \leq M$.

7. Deriving the new data set: The eigenvectors with the highest eigenvalues are projected into space, this projection results in a vector represented by fewer dimension (L < M) containing the essential coefficients.

3.3 Artificial Neural Network

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. A small neuron is a small cell that receives electrochemical stimuli from multiple sources and responds by generating electrical impulses that are transmitted to other neurons or effectors cells. There is something like 1010 to 1012 neurons in the human nervous system and each is capable of storing several bits of "information". The total weight of an average brain is 1.5kg, so an average neuron weighs something less than 1.5 X 10-9 g.

Neurons receive input from sensory or other types of cells and send outputs to other neurons or effectors organs such as muscles and glands. About 10% of neurons are input (afferent) and output (efferent). The remaining 90% are interconnected with other neurons which store information or perform various transformations on the signals being propagated through the networks. Although many different types of neurons have been identified, they all share some common characteristics. Neurons are complex cells that respond to electrochemical signals. They are composed of nucleus, a cell body, numerous dendrite links providing input connections from other neurons through synapses, and an axon trunk that carries an action potential output to other neurons through terminal links and synapses. A single neuron may be connected to hundreds or even tens of thousands of other neurons. The connections are made through two general types of synapses, excitatory and inhibitory.

Neural activity is related to the creation of an internal electric potential called a membrane potential. This potential may be increased or decreased by the input activity received from other cells through the synapses. When the cumulative inputs raise the potential above a threshold value, the neuron "fires" by propagating a sequence of action potential spikes down the axon to either excite or inhibit other neurons. The pulses cause a chemical neurotransmitter substance to be released at the terminating synapses which, in turn, can excite or inhibit other neurons. The rate of pulse propagation ranges from about 5 to 125 ms-1, and the time required for a stimulus to "traverse" a synapse is about 10ms during which the neuron cannot fire again. The activity of a neuron is measured by the firing frequency of the potential analogue spikes which it generates.

They range from about 50 to a few hundred spikes per second. Figure 3.10 below shows a simplified biological neuron.

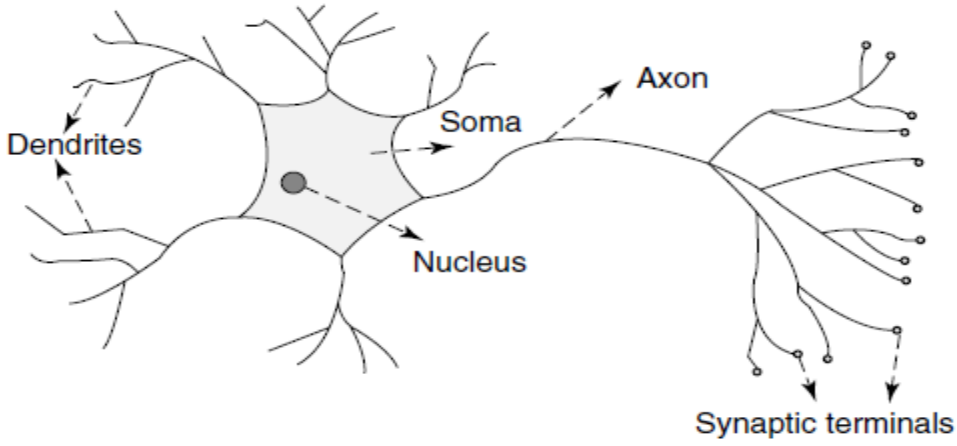


Figure 3.10: Biological neuron [23]

The above Biological neuron can be modeled as an artificial neuron shown below

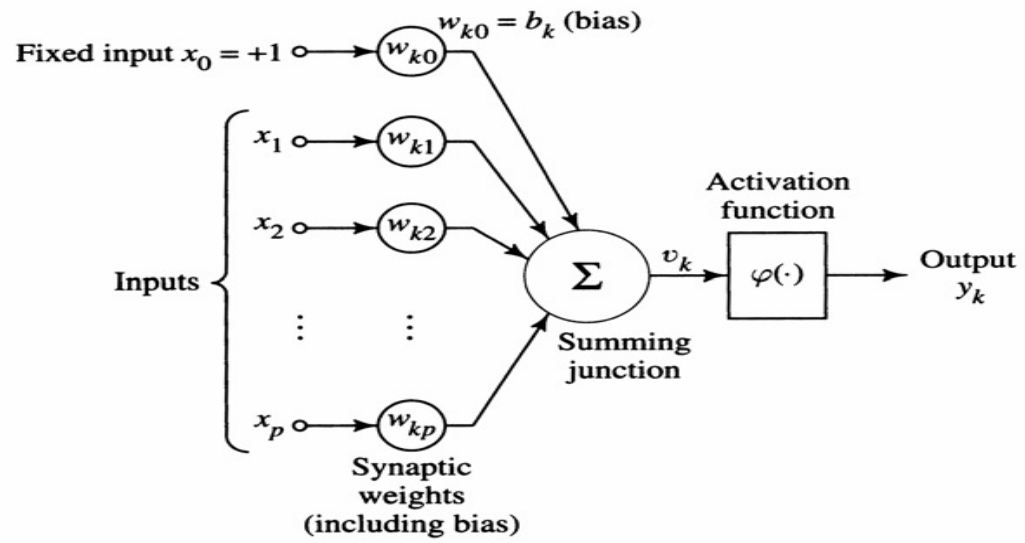


Figure 3.11: A Non linear model of a neuron as a processing device.

1. A set of synapses, each of which is characterized by a weight or strength of its own. Specifically, a signal x_j at the input of synapse j connected to neuron k is multiplied by the synaptic weight w_{kj}. The weight w_{kj} is positive if the associated synapse is excitatory; it is negative if the synapse is inhibitory.

2. An adder for summing the input signals, weighted by the respective synapses of the neuron.

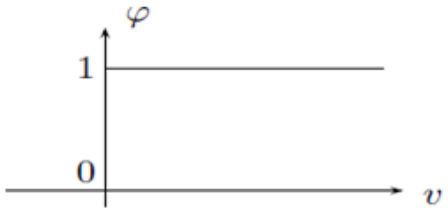
3. An activation function for limiting the amplitude of the output of a neuron. The activation function is also referred to in the literature as a squashing function in that it squashes (limits) the permissible amplitude range of the output signal to some finite value. Typically, the normalized amplitude range of the output of a neuron is written as the closed unit interval [0, 1] or alternatively [-1, 1]. The model of a neuron also includes an externally applied bias (threshold) $w_{k0} = b_k$ that has the effect of lowering or increasing the net input of the activation function.

3.3.1 Activation Function

Typically, φ(·) is a non-linear function called as activation function. Commonly used forms for φ(·) are the binary and bipolar threshold functions, the piece-wise linear function ("hard-limited" linear function), and the so-called sigmoid function. Examples of these are as follows.

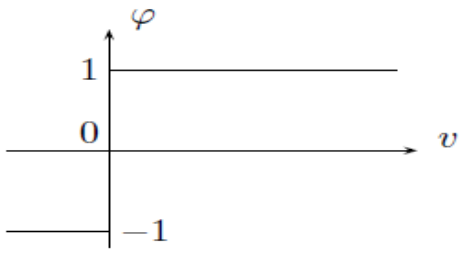
a. Binary threshold activation function:

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



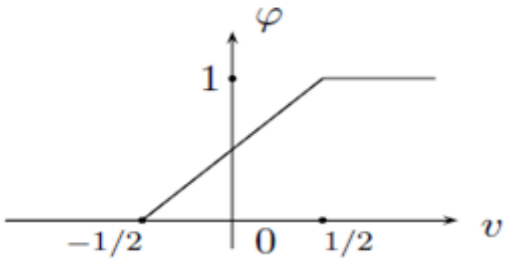
b. Bipolar threshold activation function:

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



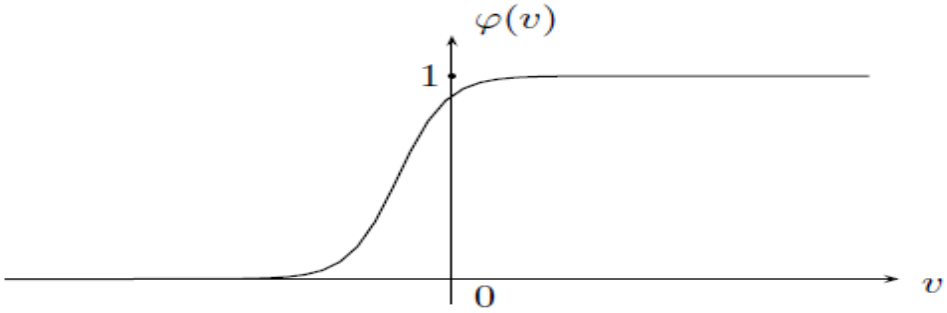
c. Piecewise-linear activation function:

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



d. Sigmoid (logistic) activation function:

A sigmoid function is any differentiable function φ (·), say, such that φ (v) → 0 as v → −∞, φ (v) → 1 as v → ∞ and φ ′(v) > 0.



A specific example of a sigmoid function is given by

$$\varphi : v \rightarrow \varphi(v) = \frac{1}{1 + \exp(-\alpha v)}$$

The larger the value of the constant parameter α, the greater is the slope. (The slope is sometimes called the "gain".) In the limit when α → ∞, this sigmoid function becomes the binary threshold function (except for the single value v = 0, for which φ (0) is equal to 1/2, for all α). One could call this the threshold limit of φ.

In mathematical terms, we may describe a neuron k by writing the following pair of equations:

$$v_k = \sum_{j=1}^{P} W_{kj} X_j \tag{3.22}$$

$$y_k = \varphi(v_k) \tag{3.23}$$

3.3.2 Neural Network Architecture

A neural network is formed when we place units at the vertices of the directed graph, with the arcs of the digraph representing the flows of signals between units. Some of the units are termed input units: these receive signals not from other units, but instead they take their signals from the outside environment. Units that do not transmit signals to other units are termed output units. The network is said to be a feed-forward network when the units can be labeled with integers in such a way that there is a connection from the computation unit labeled i to the computation unit labeled j, then i < j and it also can be in multilayer networks. This is illustrated in the Figure 3.12 below.

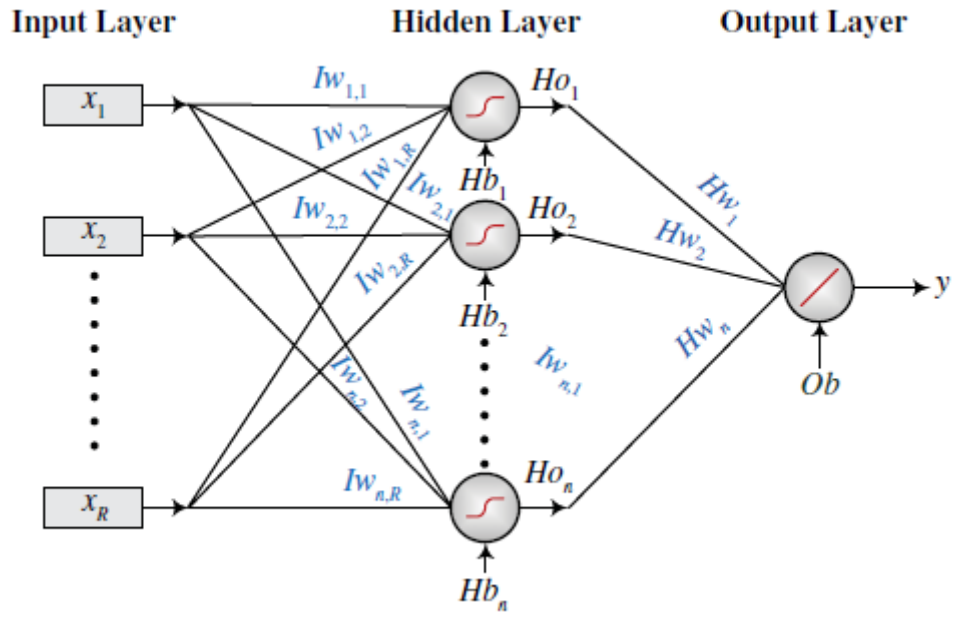


Figure 3.12: Single-hidden-layer feed forward neural network with one output. [24]

3.3.3 Neural Network Learning

Basically, there are two ways of learning in artificial neural network which are supervised and unsupervised learning. The overall concept of learning is to change the weights in the neuron in such a way that the error value is minimized for each pair of patterns of the fixed learning problem. As a stop criterion for the learning process, generally a total error E is used that is usually chosen to be the sum of the single error values. This total error is calculated a new after each epoch. An epoch is a complete run through the learning problem such that each pair of input /output patterns is processed once by the network using the learning algorithm. The learning process stops when the total error *E* is close enough to 0, or when the network is obviously not able to solve the learning problem.

A supervised learning algorithm can be divided into five steps:

i. A pair of patterns of the learning problem is chosen and the input i presented to the neural network.
ii. The input is then, propagated through the network, until it has reached its inoperative mode.
iii. The output determined by the network is compared to the target pattern. The error value (e) is determined and added to the total error E.
iv. If $e \neq 0$, then the weights are changed in a way that a reduction of the absolute value of error.

The most common supervised learning algorithms involve approximation of a gradient descent method and try to reduce the total error E to zero. However, this learning procedure cannot always guarantee convergence, because it is equivalent to a local heuristic search procedure. The algorithm stops when it reaches a local minimum. There are however, some disadvantages of this technique as the local minimum is not the global minimum. This means that the minimized error is not reached and the learning process has failed.

On the other hand, unsupervised learning algorithm works on a free learning problem which only contains input patterns. However, for this kind of learning procedure the network is also supposed to map similar input patterns to output that are similar to each other. In unsupervised learning algorithm, the units compete with each other and the principle of winner-takes-all applied. It is trained without teaching signals or targets and it is only supplied with examples of

the input patterns that it will solve eventually. This type of learning usually has an auxiliary cost function which needs to be minimized and the weights are modified where a cost function is minimized. At the end of the learning phase, the weights would have been adapted in such a manner such that similar patterns are clustered into a particular node.

3.3.3.1 Perceptron learning rule:

The perceptron is the simplest form of a neural network used for the classification of a special type of patterns, which are linearly separable. It consists of a single McCulloch-Pitts neuron with adjustable synaptic weights and bias (threshold). Rosenblatt proved that if the patterns (vectors) used to train the perceptron are drawn from linearly separable classes, then the perceptron algorithm converges and positions the decision surface in the form of a hyperplane between the classes. The proof of convergence of the algorithm is known as the perceptron convergence theorem. The single-layer perceptron shown has a single neuron. Such a perceptron is limited to performing pattern classification with only two classes.

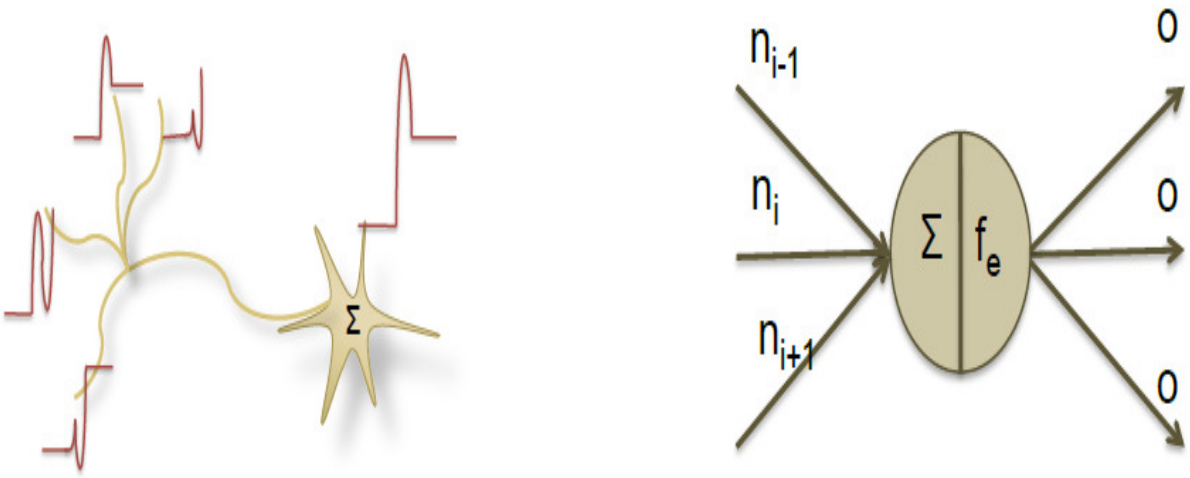


Figure 3.13: Perceptron (left), computing unit of ANN, analogous to a biological neuron (right) [25]

The training technique used is called the perceptron-learning rule. Perceptrons are especially suited for simple problems in pattern classification. Suppose we have a set of learning samples consisting of an input vector x and a desired output d(k). For a classification task, the d(k) is usually +1 or −1. The perceptron-learning rule is very simple and can be stated as follows:

1. Start with random weights for the connections.
2. Select an input vector x from the set of training samples.
3. If output $y_k \neq d(k)$ (the perceptron gives an incorrect response), modify all connections w_i according to:

$$\delta w_i = \eta\,(d_k - y_k)\,x_i\ ;\ (\eta = \text{learning rate}).$$

$$w_i\,(\text{new}) = w_i(\text{old}) + \eta\,(d_k - y_k)\,x_i\ ;\ \ w_i(\text{new}) \text{ is updated value of weight}$$

4. Go back to step 2.

3.3.4 Back propagation Learning:

Multilayer perceptrons have been applied successfully to solve some difficult diverse problems by training them in a supervised manner with a highly popular learning known as the error back-propagation learning. This is based on the error-correction learning rule. Basically, the error back-propagation process consists of two passes through the different layers of the network: a forward pass and a backward pass. In the forward pass, activity pattern (input vector) is applied to the sensory nodes of the network, and its effect propagates through the network, layer by layer. Finally, a set of outputs is produced as the actual response of the network. During the forward pass the synaptic weights of network are all fixed. During the backward pass, on the other hand, the synaptic weights are all adjusted in accordance with the perceptrons error-correction rule as discussed above. Specifically, the actual response of the network is subtracted from a desired (target) response to produce an error signal. This error signal is then propagated backward through the network, against direction of synaptic connections - hence the name “error back-propagation”. The synaptic weights are adjusted so as to make the actual response of the network move closer the desired response.

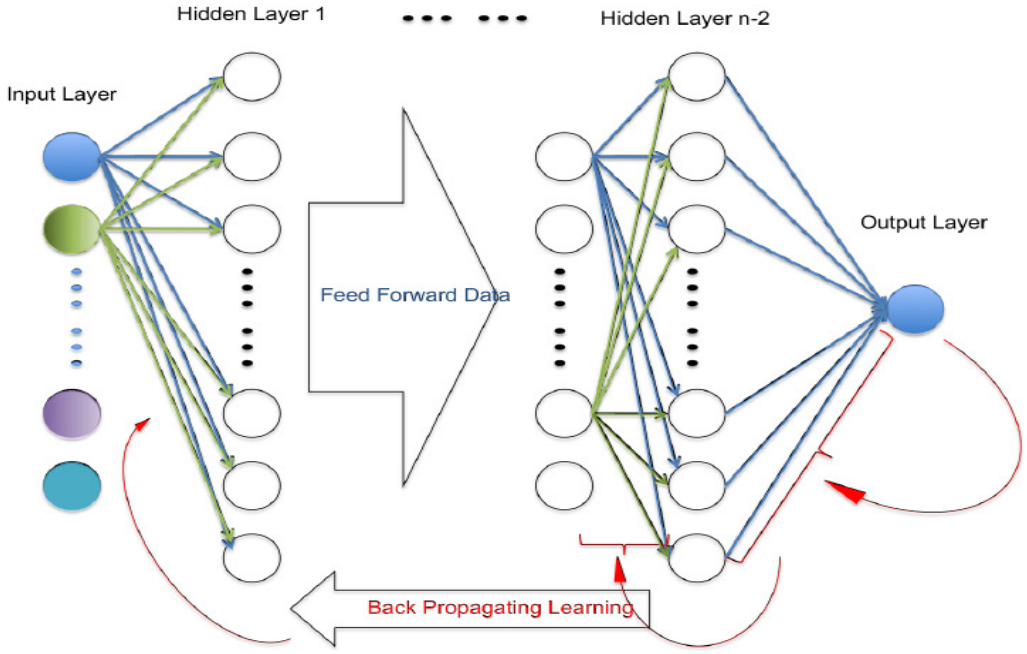


Figure 3.14: A Feed-forward network with back propagation learning [25]

The feed-forward back-propagation neural network is shown in Figure 3.14. It is fully connected, which means that a neuron in any layer is connected to all neurons in the previous layer. Input data owes the direction from left to right side (Feed Forward Data arrow) and learning process goes on left-to-right direction to adjust nodes weights to improve next trial (Back Propagating Learning arrow). Signal flow through the network progresses in a forward direction, from left to right and on a layer-by layer basis.

3.3.5 Multi-class pattern classification using neural networks

Multi-class pattern recognition is a problem of building a system that accurately maps an input feature space to an output space of more than two pattern classes. While two class classification problem is well understood, multi-class classification is relatively less investigated. Many pattern classification systems were developed for two-class classification problems and theoretical studies of learning have focused almost entirely on learning binary functions [26] including artificial neural network algorithms such as the perceptron and the error back propagation (BP) algorithm[27][28][29].

3.3.5.1 An overview of neural network systems for multi-class pattern classification

A K-class pattern classification problem can be implemented in either one of the two neural network architectures, a single neural network system with M outputs, where M>1 (see Fig. 3.15(a)) or a system of multiple neural networks (see Fig. 3.15(b) and (c)).

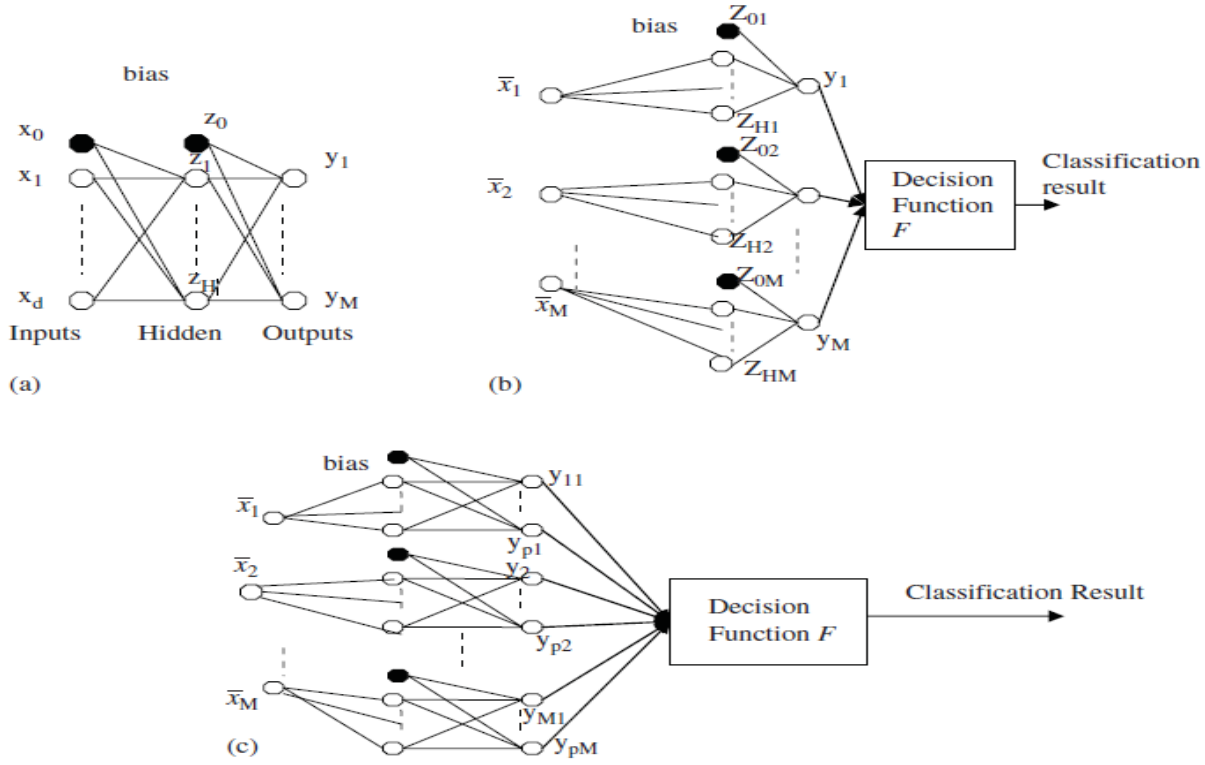


Figure 3.15: Different neural network architectures for implementing K-class pattern classification

Fig. 3.15(a) shows a single neural network for pattern classification of K-classes. The number of the output nodes M, is determined by the encoding scheme for pattern classes, and is not necessary equal to K. Fig. 3.15(b) illustrates a system of M binary neural networks used to classify K object classes with a decision module that integrates the results from the M binary neural networks, and Fig. 3.15(c) illustrates a system of multiple neural networks for multi-class pattern classification each with multiple output nodes. Note the feature vectors in different neural networks in Fig. 3.15(b) and (c) can be different from each other.

Hence, the complete analysis of Brain MRI Image classification can be explained in term of following schematic diagram shown below.

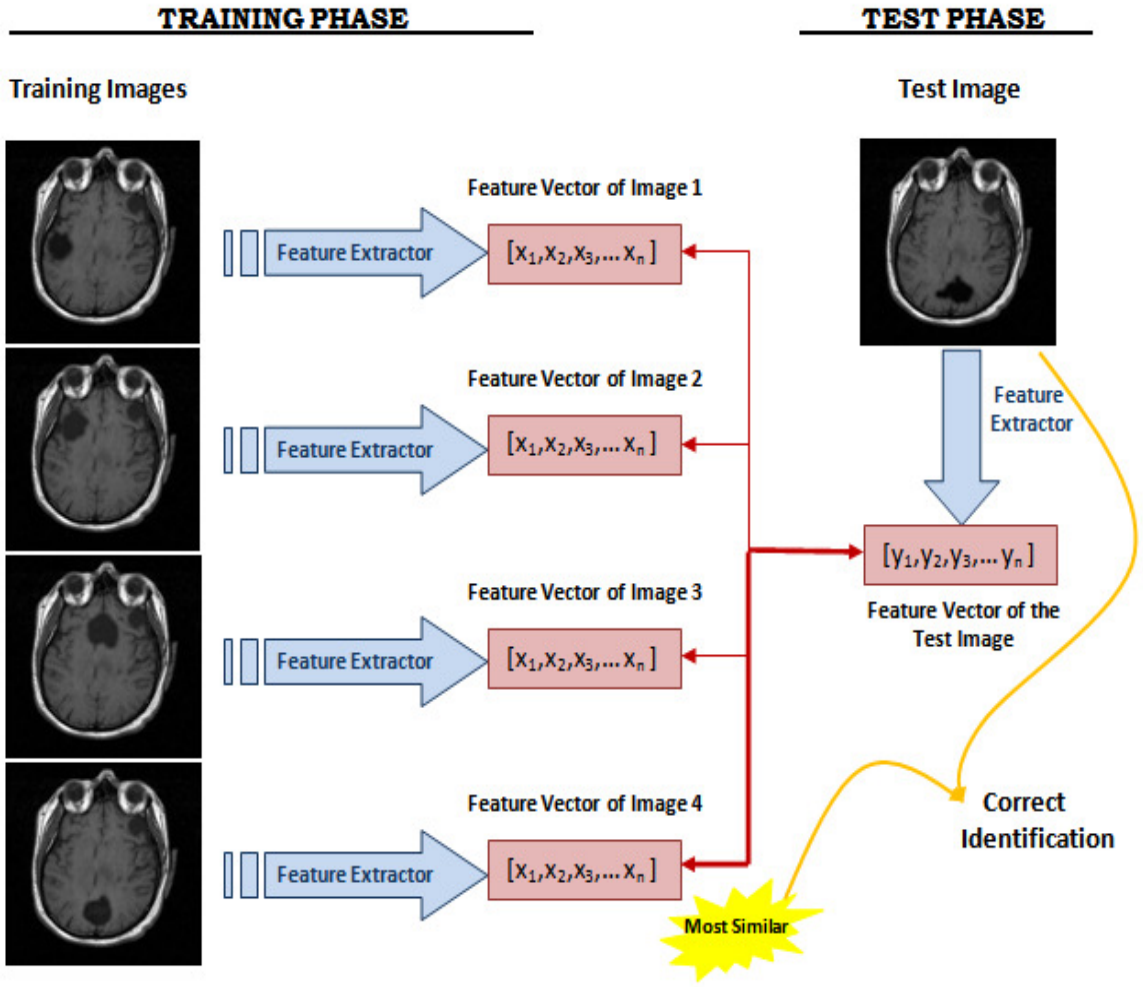


Figure 3.16: Schematic Diagram of Brain MRI classification overview [30]

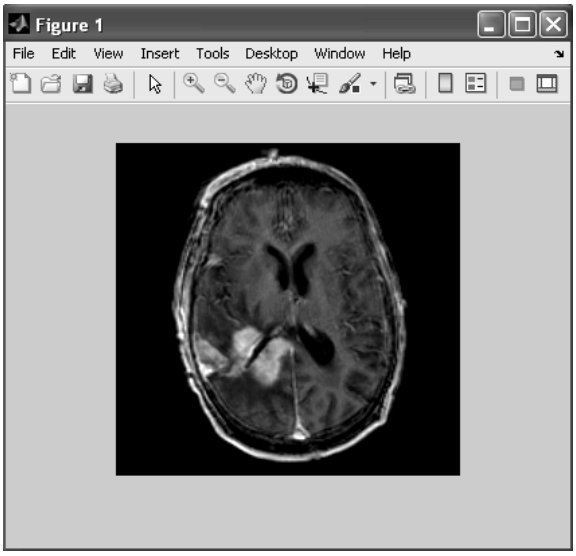
CHAPTER 4

SIMULATION RESULTS

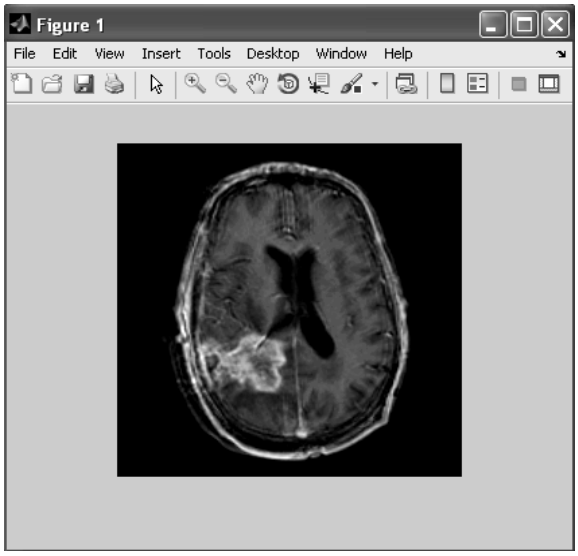
Following sets of MRI data are taken for training & testing purpose.

TRAINING DATA

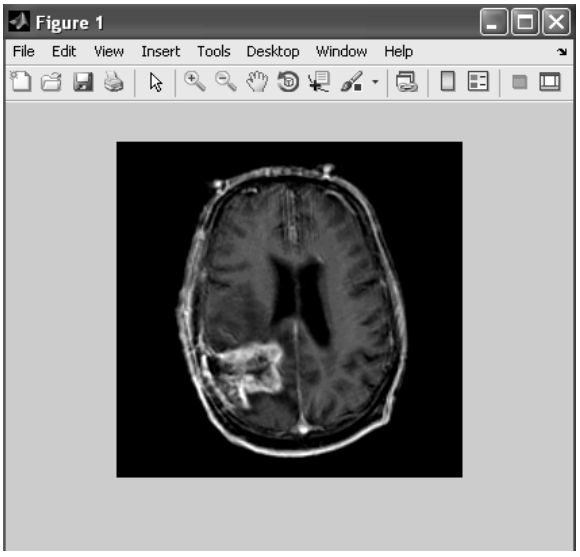
CLASS 1:



(a).



(b).



(c).

CLASS 2:

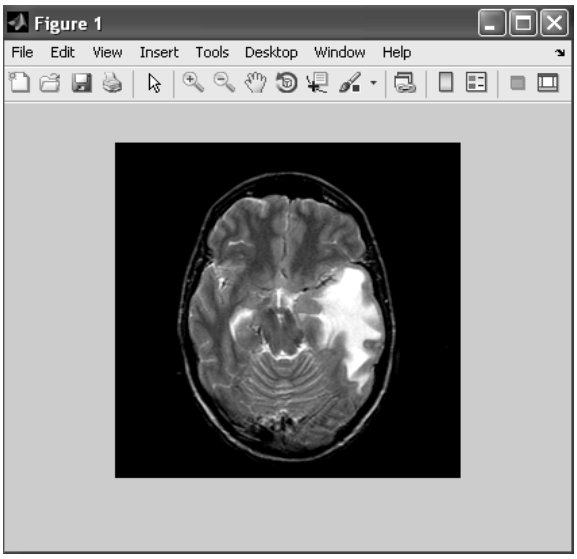


Figure 1
File Edit View Insert Tools Desktop Window Help

(d).

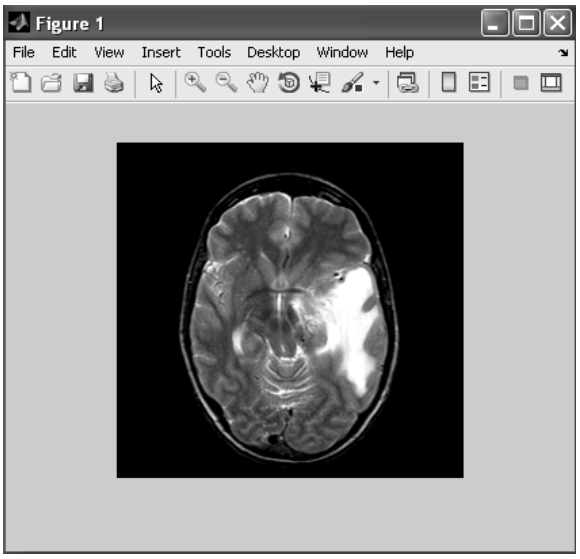


Figure 1
File Edit View Insert Tools Desktop Window Help

(e).

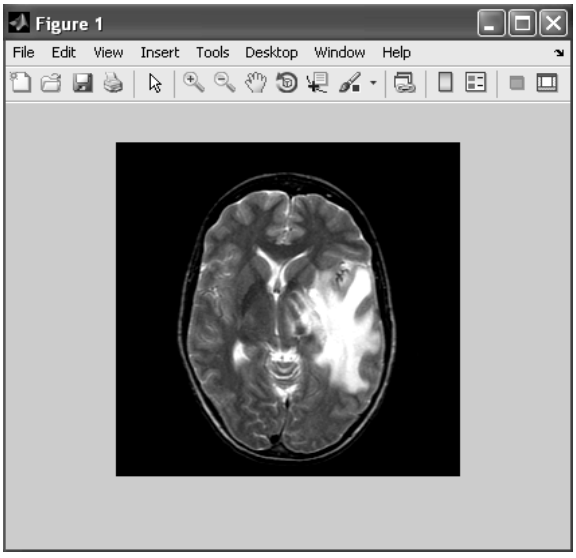


Figure 1
File Edit View Insert Tools Desktop Window Help

(f).

CLASS 3:


Figure 1
File Edit View Insert Tools Desktop Window Help

(g).

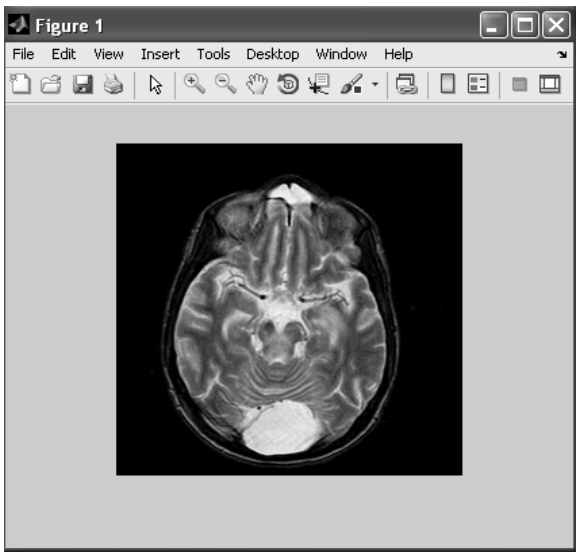


Figure 1
File Edit View Insert Tools Desktop Window Help

(h).

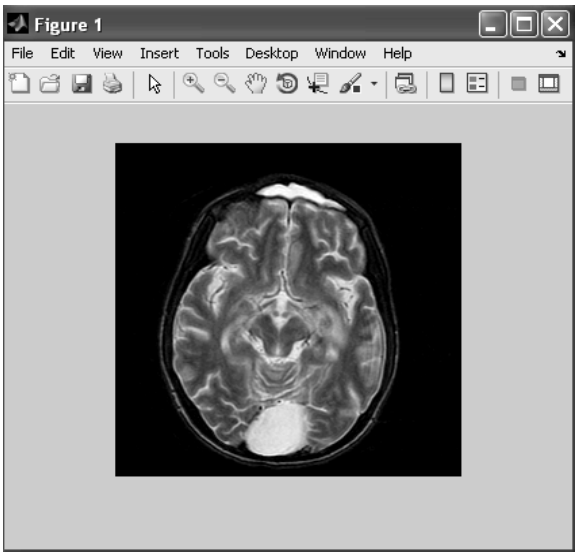


Figure 1
File Edit View Insert Tools Desktop Window Help

(i).

CLASS 4:

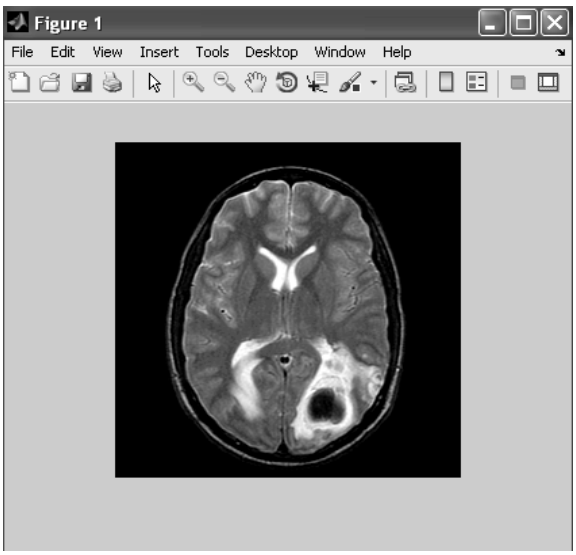


Figure 1
File Edit View Insert Tools Desktop Window Help

(j).

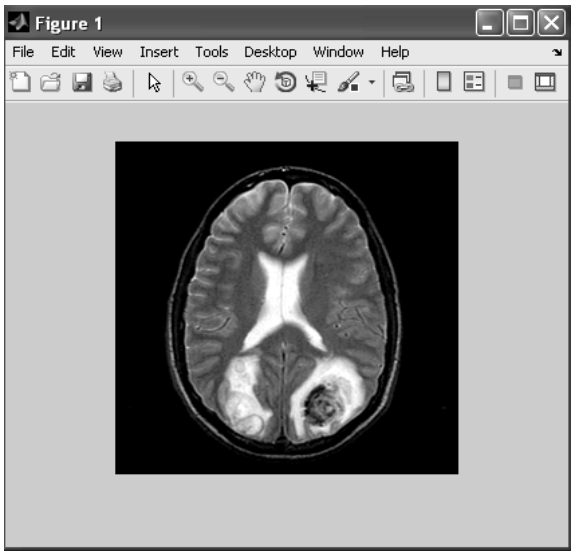


Figure 1
File Edit View Insert Tools Desktop Window Help

(k).

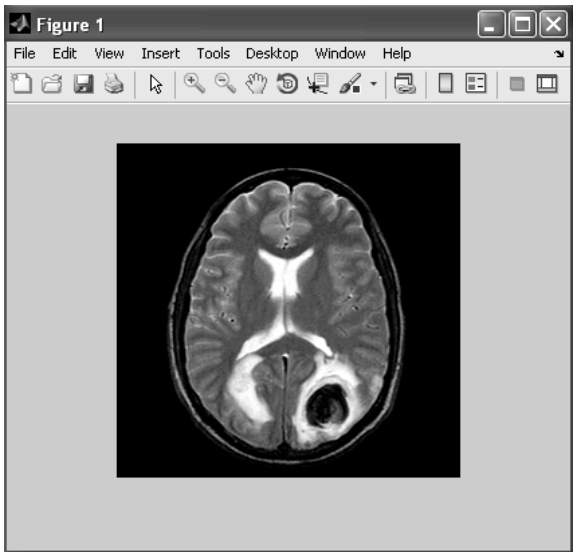
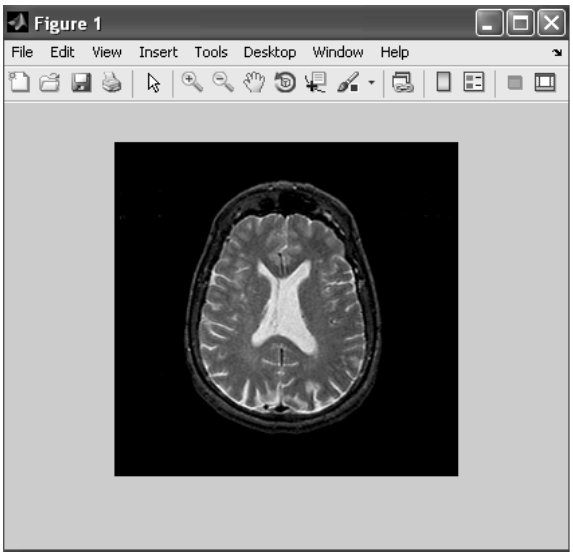


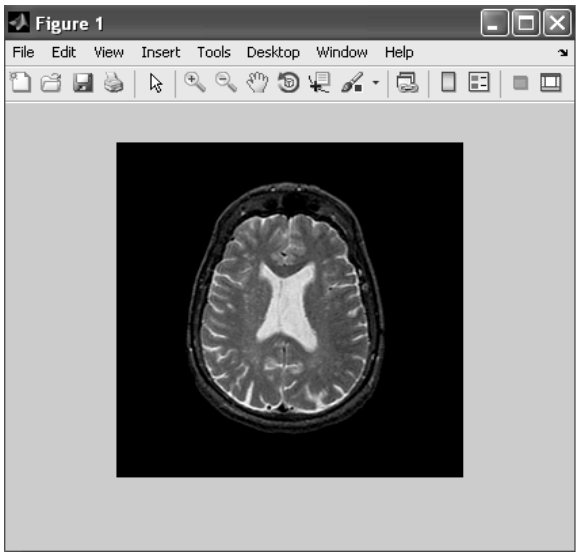
Figure 1
File Edit View Insert Tools Desktop Window Help

(l).

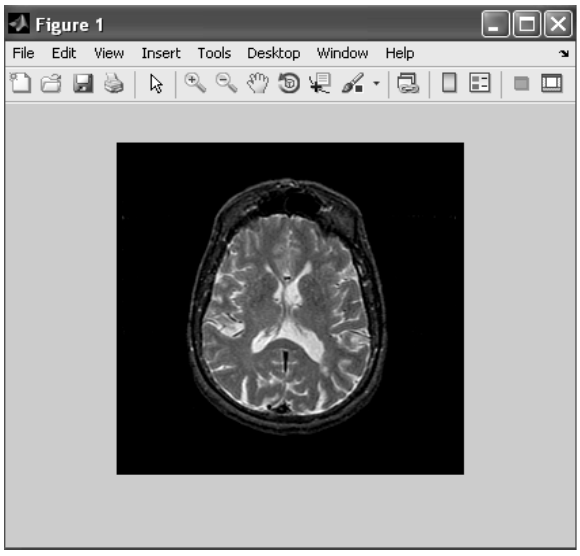
CLASS 5:



(m).



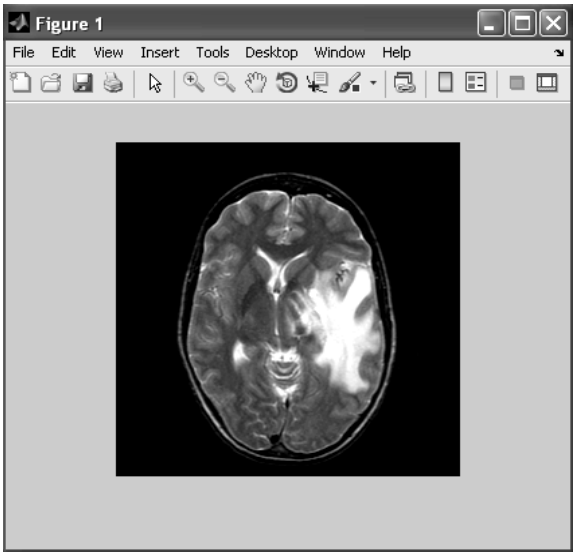
(n).



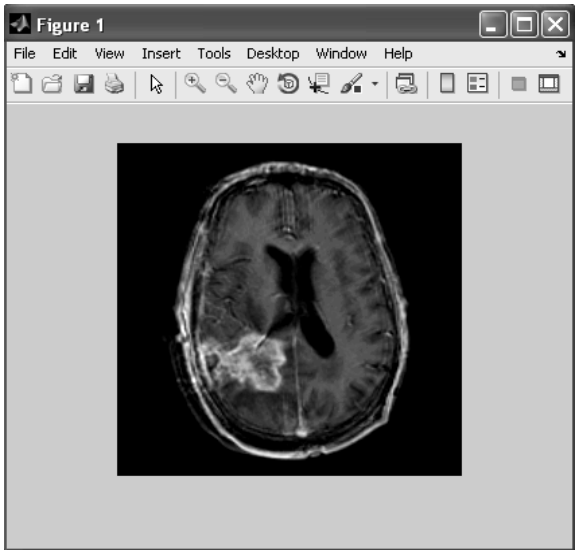
(o).

Figure: 4.1 Training data set

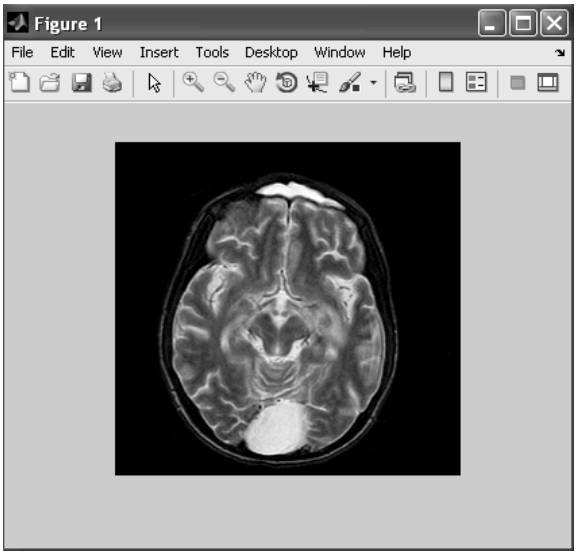
TEST DATA



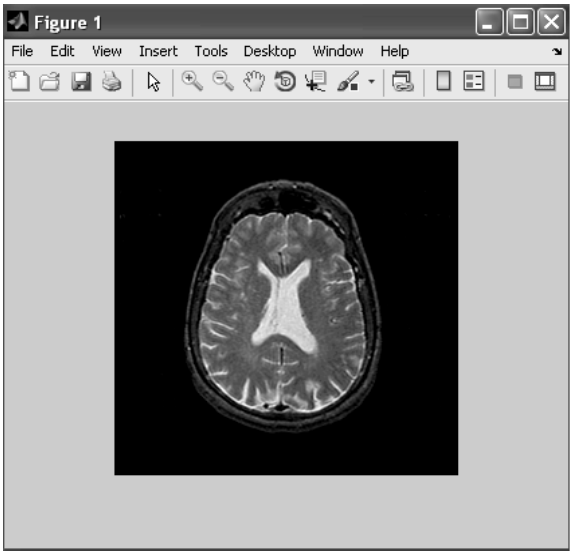
(a).



(b).



(c).



(d).

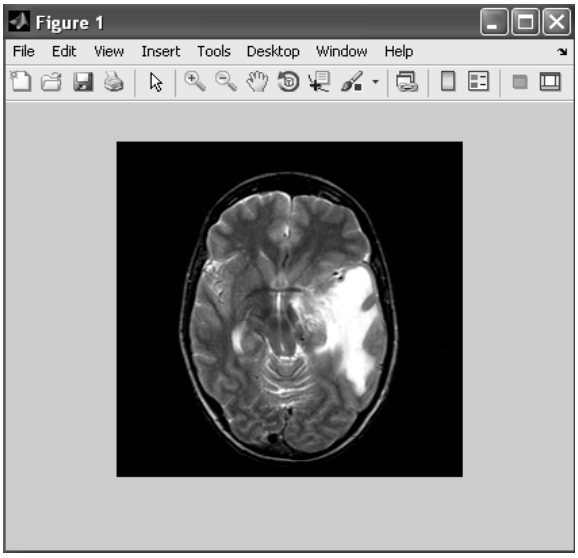


Figure 1
File Edit View Insert Tools Desktop Window Help

(e).

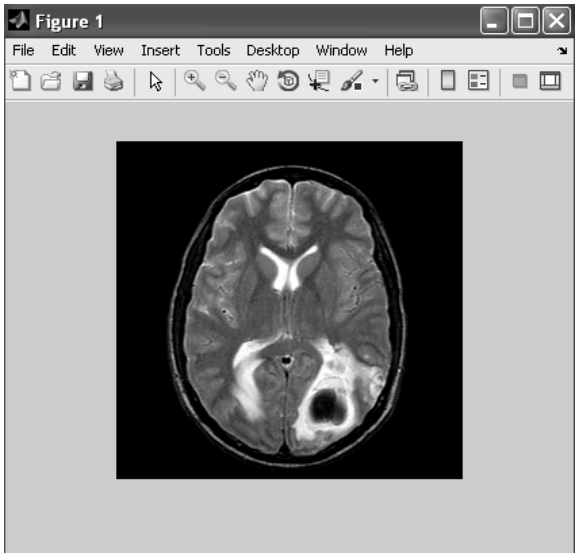


Figure 1
File Edit View Insert Tools Desktop Window Help

(f).

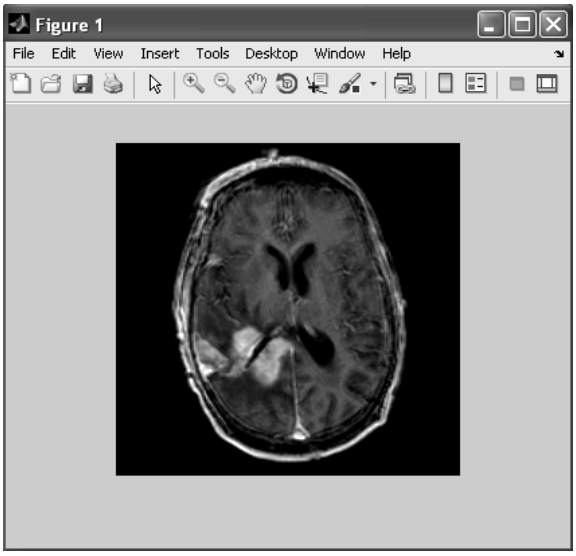


Figure 1
File Edit View Insert Tools Desktop Window Help

(g).

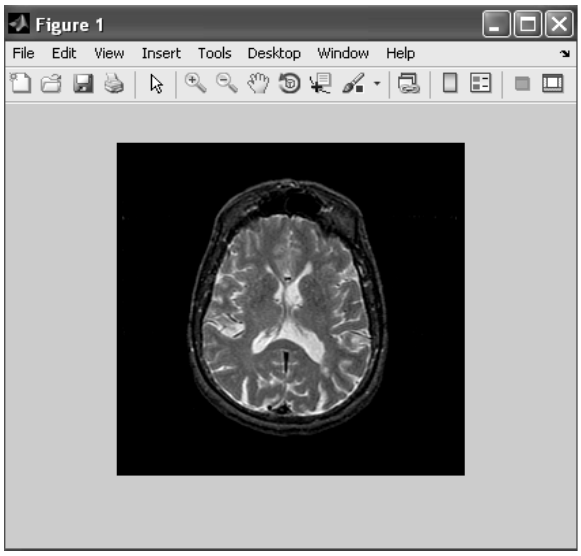


Figure 1
File Edit View Insert Tools Desktop Window Help

(h).

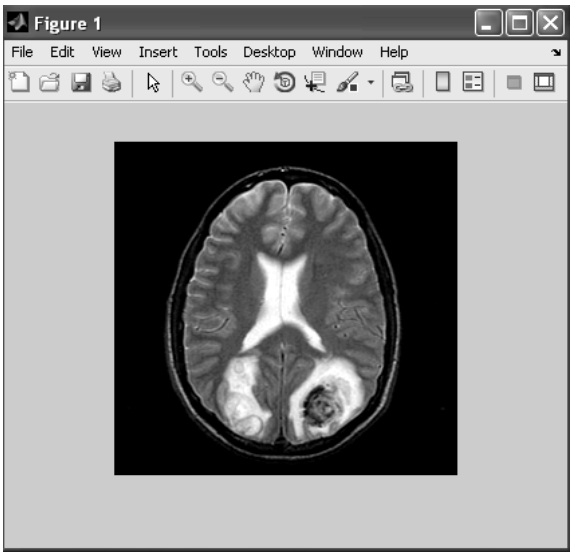


Figure 1
File Edit View Insert Tools Desktop Window Help

(i).


Figure 1
File Edit View Insert Tools Desktop Window Help

(j).

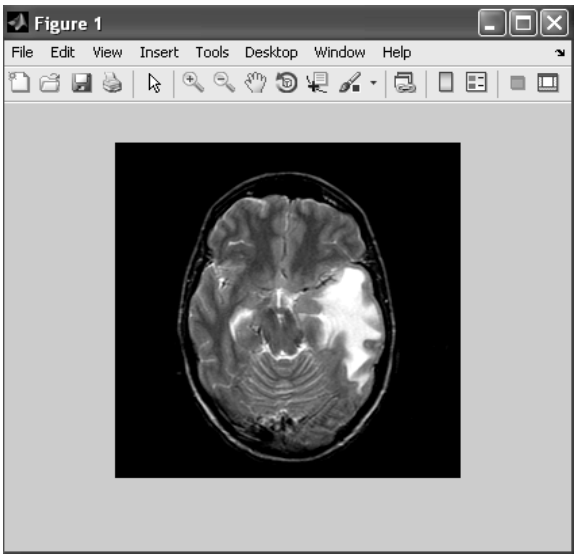


Figure 1
File Edit View Insert Tools Desktop Window Help

(k).

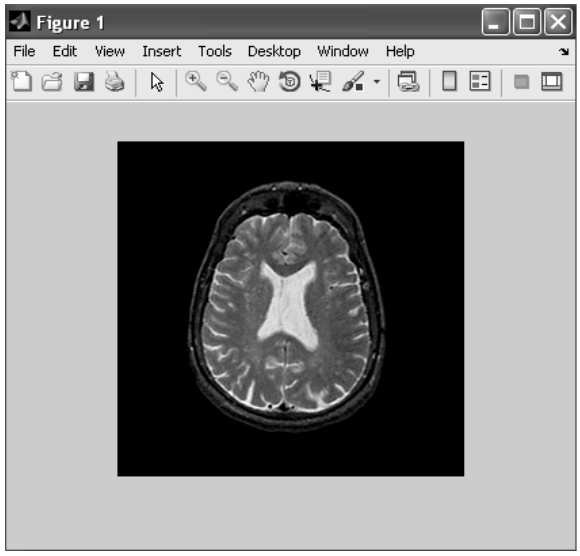
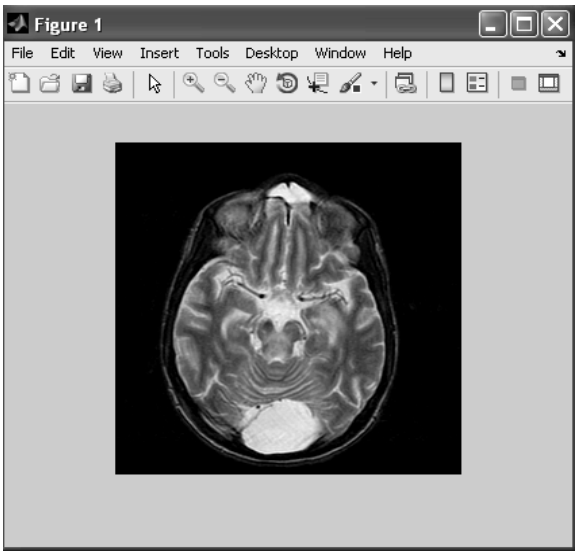
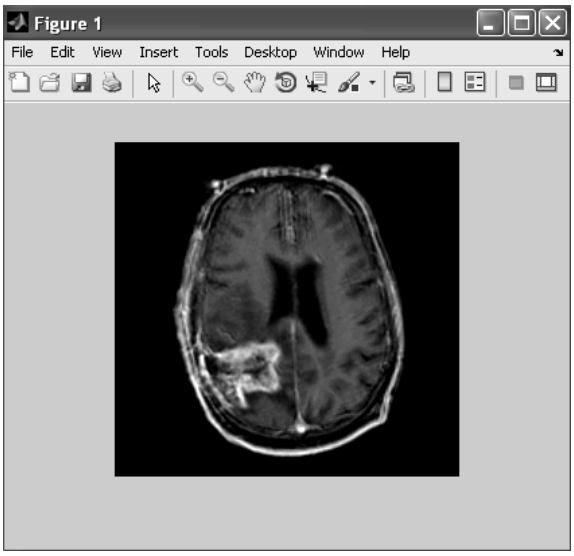


Figure 1
File Edit View Insert Tools Desktop Window Help

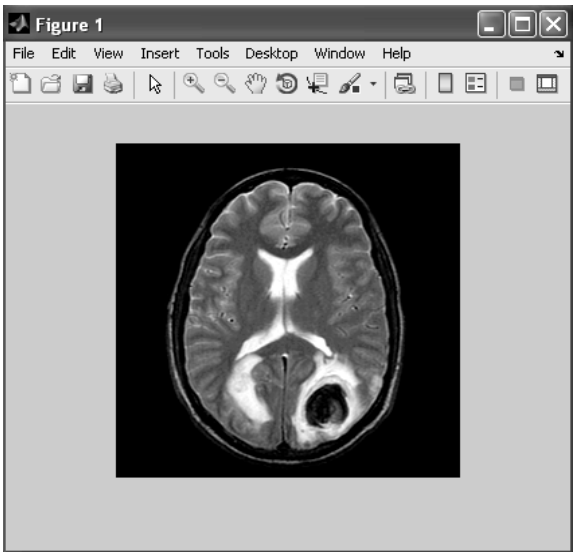
(l).



(m).



(n).



(o).

Figure 4.2: Testing data set

Five sets of Training data is taken. Each set contains three Images .One set contains normal MRI Images & rest four sets contain abnormal or brain tumor infected MRI images. The abnormal MRI images in four sets contain tumor location in different part of brain. Testing is done on fifteen MRI images. The classification is done using artificial neural network & is implemented on MATLAB. The results are obtained in the form of confusion matrix as shown below.

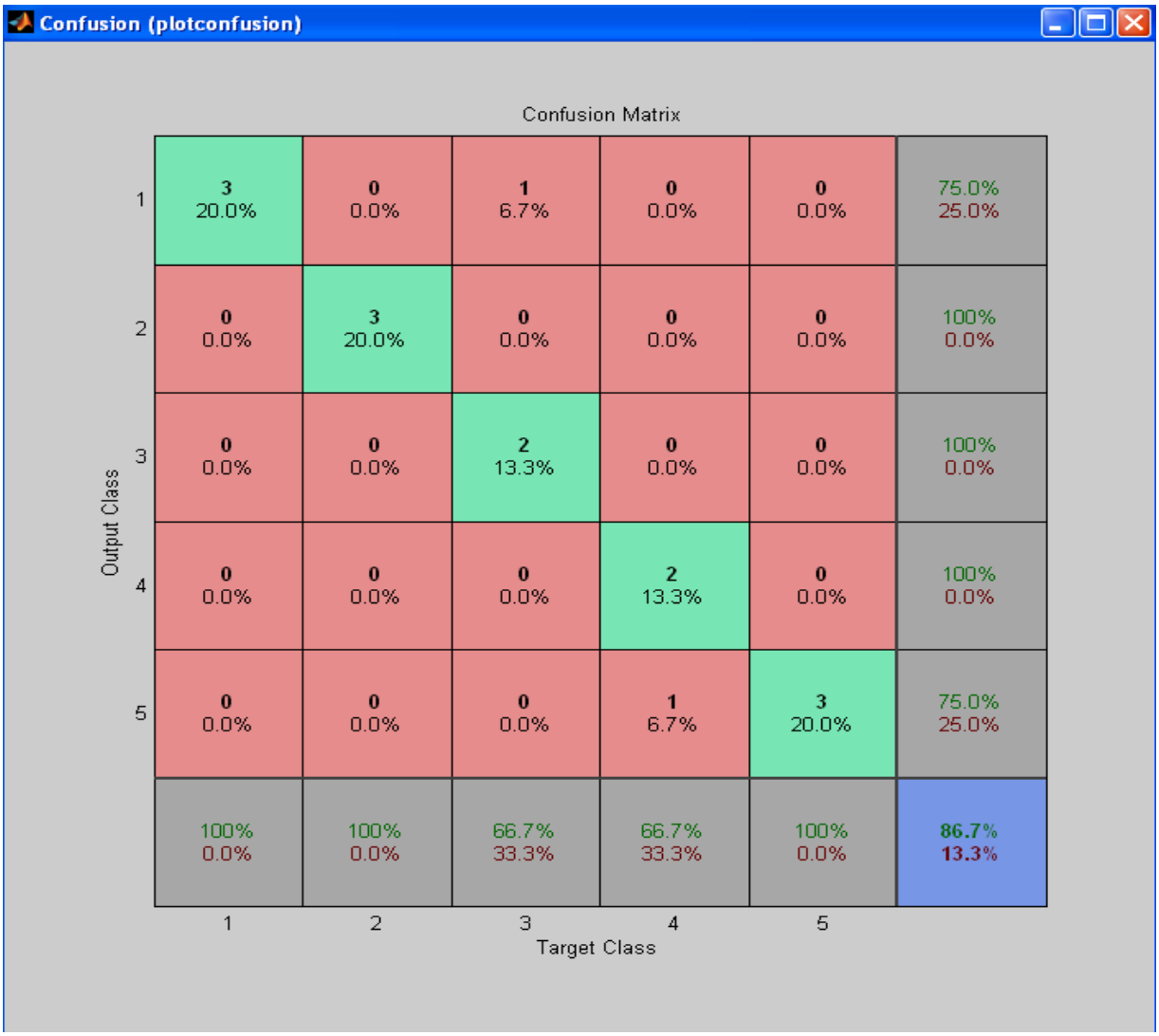


Figure 4.3: Confusion Matrix

Matrix shows that classification using artificial neural network gives 86.7% accurate results which are very good results as here number of training data sets taken are very less. If large training data is taken then results would be having excellent accuracy.

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

This work proposed the development of an automated brain MRI diagnostic system, which can classify whether the MR image belongs to a normal brain or to a person suffering from brain tumor & also find out the location of tumor in the brain . For this purpose, five classes of training data have been taken. The scheme employs a three -stage algorithm. Firstly, discrete wavelet transform using Haar wavelet is used for extracting features vectors. Vectors computed here are wavelets coefficients matrices of M.R.I Images of different training data sets. Then Principal component analysis is applied to get dominant features vectors & thus reducing the number of features vectors. Finally, for classification artificial neural network is used. Here, firstly neural network is trained using training data. Once, the neural network is trained classification is done on test data. Confusion matrix obtained by classification give 86.7% accuracy which are very good results as here training data is very less. Higher accuracy can be obtained using very large training data set.

5.2 Future Scope

Here, discrete wavelet transform is used as feature extraction technique. Although, latest research on brain MRI analysis state that DWT gives best features vectors but due to various others transforms techniques available for extracting image features DWT can be replaced by other method. So future scope of current work is adopting different method of feature extraction so that higher classification rates can be obtained.

BIBLIOGRAPHY

[1]. Brain Tumor Facts & Statistics: http://www.sdbtf.org/facts-bout-bt.html

[2]. Cancer mortality in India: a nationally representative survey
http://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2812%2960358-4/abstract

[3]. Neeraj Sharma, Lalit M. Aggarwal, "Automated medical image segmentation techniques", J. Med. Physics. Vol. 35, 3-14, 2010.

[4]. Yu-hua Chai, Li-qun Gao, Shun Lu and Lei Tian, "Wavelet-based Watershed for Image Segmentation Algorithm" , Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21 - 23, 2006, Dalian, China.

[5]. M. Sonka, J. M. Fitzpatrick: "Handbook of Medical Imaging - Processing and Analysis", I.N. IEEE Trans. Med. Imaging 20(3): 249-250 (2001).

[6]. O. Salvado, P. Bourgeat, O.A. Tamayo, M. Zuluaga, S. Ourselin, "Fuzzy classification of brain MRI using a priori knowledge: Weighted fuzzy C-means", Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference, pp 1-8, 2007.

[7]. Horowitz, A. L. (1995), "MRI physics for radiologists", (3rd ed.) New York: Springer-Verlag.

[8]. Buxton, R. B. (2002), "Introduction to functional magnetic resonance imaging: Principles and techniques", Cambridge: Cambridge University Press.

[9]. Hashemi, R. H., Bradley, W. G., & Lisanti, C. J. (2004), "MRI the basics", (2nd ed.) Philadelphia: Lippincott Williams & Wilkins.

[10]. I. Daubechies, "Ten lectures on wavelets", vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1992.

[11]. Xi Yang, Dawei Qi, Xianhong Li, "Multi-scale Edge Detection of Wood Defect Images Based on the Dyadic Wavelet Transform", 2010 International Conference on Machine Vision and Human-machine Interface.

[12]. G. Strang and T. Nguyen, "Wavelets and Filter Banks", Wellesley-Cambridge Press, second edition, 1997. ISBN 0-9614088-7-1.

[13]. M.G.E. Schneiders, "Wavelets in control engineering", Master's thesis, Eindhoven University of Technology, August 2001. DCT nr. 2001.38.

[14]. G Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", IEEE transactions on pattern analysis and machine intelligence, VOL II. No 7 11(7):674-693 Juillet 1989.

[15]. Mamta Juneja, Parvinder Singh Sandhu, "Performance Evaluation of Edge Detection Techniques for Images in Spatial Domain", International Journal of Computer Theory and Engineering, Vol. 1, No. 5, December, 2009 1793-8201.

[16]. Jun Tang , "A Color Image Segmentation algorithm Based on Region Growing" , School of Electronic Engineering ,Xi'an Shiyou University ,Xi'an ,China.

[17]. Stephane Mallat and Sifen Zhong, "Signal Characterization From Multiscale Edges", Courant Institute of Mathematical Sciences New York University.

[18]. R. C. Gonzalez and R. E. Woods, "Digital Image Processing," 2nd edition, New Jersey, Prentice Hall, 2008.

[19]. K. Roy, and P. Bhattacharya, "Optimal features subset selection and classification for Iris recognition", Journal of Image Video Process. , Mar. 2008.

[20]. N.Hema Rajini, R.Bhavani ,"Classification of MRI Brain Images using k- Nearest Neighbor and Artificial Neural Network", IEEE-International Conference on Recent Trends in Information Technology, ICRTIT 2011, MIT, Anna University, Chennai. June 3-5, 2011.

[21]. A. K. Jain, P. W. Robert Duin, and M. Jianchang, "Statistical pattern recognition: A review," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, pp.4–37, Jan.2000.

[22]. Jonathon Shlens , "A Tutorial on Principal Component Analysis" , Center for Neural Science, New York University New York City, NY 10003-6603 and Systems Neurobiology Laboratory, Salk Institute for Biological Studies La Jolla, CA 92037 , Electronic address: shlens@salk.edu

[23]. "Artificial Neural Networks", Ajith Abraham, Oklahoma State University, Stillwater, OK, USA, Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn. ©2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

[24]. Saman Razavi and Bryan A. Tolson , "A New Formulation for Feed forward Neural Networks" , IEEE Transactions on Neural Networks, Vol. 22, No. 10, October 2011 , Digital Object Identifier 10.1109/TNN.2011.2163169.

[25]. Eun Young Kim , "Multistructure segmentation of multimodal brain images using artificial neural networks." thesis, Iowa Research Online, University of Iowa, 2009. http://ir.uiowa.edu/etd/387.

[26]. B.K. Natarajan, "Machine Learning: A Theoretical Approach", Morgan Kaufmann, Los Alamitos, CA, 1991.

[27]. C.M. Bishop, "Neural Networks for Pattern Recognition", Oxford University Press, Oxford, 1995.

[28]. D. Rumelhart, G. Hinton, R. Williams, "Learning internal representations by error propagation", : D. E. Rumelhart, J. L. McClelland (Eds.), Parallel Distributed Processing: Explorations in the Microstructures of Cognition, vol. 1, MIT Press, Cambridge, MA, 1986, pp. 318–362.

[29]. M. Anthony, P. Bartlett, "Neural Network Learning: Theoretical Foundations", Cambridge University Press, Cambridge, UK, 1999.

[30]. Mohd Fauzi Othman and Mohd Ariffanan Mohd Basri, "Probabilistic Neural Network for Brain Tumor Classification", 2011 Second International Conference on Intelligent Systems, Modelling and Simulation, 978-0-7695-4336-9/11, © 2011 IEEE.

[31]. Harvard Medical School, Web, data available at http://med.harvard.edu/AANLIB/.