Malignant Cancer Cell Detection By Microscopy Image Analysis

BY

Md. Najmul Hussain
ID: 102-15-1059

AND

Mohammad Azizur Rahman

ID: 102-15-1060

AND

Tomal Mahdi

ID: 102-15-1062

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mohammad Nazmul Haque
Senior Lecturer

Along with
Mohammad Samiullah
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

OCTOBER 2012

APPROVAL

This Project titled **"Malignant Cancer Cell Detection By Microscopic Image Analysis**", subsmitted by Md. Najmul Hussain, Mohammad Azizur Rahman and Tomal Mahdi to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on October 18, 2012.

BOARD OF EXAMINERS

_____

Dr Syed Akhter Hossain **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

_____

Dr Yousuf Mahbubul Islam **Internal Examiner**
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

_____

Dr Md Kabirul Islam **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

_____

Dr Muhammad Shorif Uddin **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of Mohammad Nazmul Haque**, Senior Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

_____

Mohammad Nazmul Haque
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

_____

Md. Najmul Hussain
ID: 102-15-1059
Department of CSE
Daffodil International University

_____

Mohammad Azizur Rahman
ID: 102-15-1060
Department of CSE
Daffodil International University

_____

Tomal Mahdi
ID: 102-15-1062
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty Allah for His divine blessing makes us possible to complete this project successfully.

We fell grateful to and wish our profound our indebtedness to **Mohammad Nazmul Haque**, **Senior Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Image Processing influenced us to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Mohammad Nazmul Haque, Mohammad Samiullah, and Dr Syed Akhter Hossain, Professor and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

This project is on "**Malignant Cancer Cell Detection By Microscopy Image Analysis**". This is a kind of image processing work, which help to check the malignancy of a cell by testing a biopsy image and provide faster salutation than traditional diagnosis system.

The aim of the system is to check malignancy of a cells biopsy image for this firstly we do color based segmentation. After Color based segmentation we have done Clustering by using k-means algorithm. When clustering is done we have selected the best cluster among them and do feature extraction like nucleus counting and number of white and black pixel. After that we compared it weather it is in the ratio of class normal, invasive or insitu which indicate its malignancy level.

To develop this project the most essential tool was MATLAB which helps to simulate all the parts of the work and test it. Another tool that is used in the development work is Java technology. Some other tools as photoshop, Jaspersoft for iReport are used in some portions of the development.

After implementation of all functions, the system is tested in different stages and it works successfully as a prototype.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

CHAPTER 1: INTRODUCTION

Microscopic Image is one special type of image which is taken with the help of microscope which is mainly used for different aspects of the human body diagnosis and cell analysis. And Microscope image processing is a process that uses the technique of digital image processing for processing, analyzing and recognizing microscopic image. This process is now used in a number of various fields such as medicine, biological research, cancer research, drug testing, metallurgy, etc. As for this cause we are using microscopic image for our project.

Cancer is one of the most dangerous or life threatens disease throughout the world. A statistics during this year estimated that 1,638,910 men and women (848,170 men and 790,740 women) will be diagnosed and 577,190 men and women will die for cancer of all sites [6]. A survey throughout our country shows that per year 17686 women over 100,000 women affected for cervical cancer [7] and A survey done in 2001 showed that 22000 women were affected every year by breast cancer and 17000 (77%) of them died. This rate is because the manual technique for cancer detection is a lengthy process, as for why people are dying of this disease. As for this we are working for malignant cancer detection using microscopic image.

Malignant Cancer Cell Detection by Microscopic Image Analysis is a process for finding out the malignancy level of a cell or weather the cell is cancer affected or not. This includes some process: Preprocessing, Segmentation and feature extraction and Decision making.

There are many works already done in this topic. We analyze some of them. In most of the work they are working like, after converting to binary image them doing segmentation, clustering and other process. They are not working with color image. There also weakness in the feature extracting process. There are some work which needed to process the cell before the image capturing which is time consuming and inefficient.

1.1 Motivation

A huge number of affected people are dying because of breast cancer. To reduce the number of death and cure the people the first and foremost way is to detect the cancer immediately that means when it is in the first stage of the disease. If it is possible to do it then it can be cured. But the traditional manual detection process is time consuming. This is why research is happening on how to detect breast cancer. The most interesting and possible field is diagnosis by microscopic image analysis of the cells.

We are motivated in the topic of cancer detection by microscopic image analysis because for cancer detection through image analysis requires image processing and different data mining or neural network. Although many a works are done in this arena and achieved great accuracy, we are want work in this field because we are trying to propose a system with great accuracy and proficiency in a different way. Most of the work are weather emphasize on noise removal or feature extraction. But we have given importance in both noise removal and feature extraction. And then on the base of the features we would take a decision.

1.2 Objective

The main objective of this project is to develop a system that will be able to detect cancer by analyzing a biopsy image a cell. For developing such a system we followed some steps including color based segmentation, noise removal, feature extraction and decision making by data mining. In this process we use color based segmentation for finding best feature extraction which help us in doing proper and more accurate decision making. Though already many systems for cancer detection by testing biopsy image, our **main objective** of doing this work is develop a system that will detect cancer cell through **simple processing** of the image and get **more accuracy** than previous system.

1.3 Contribution

We have proposed a system that contains color based segmentation and then clustering based on the segmentation which is different from other works on this field of cancer detection. We are using this technique because the segmentation provides easier way to

differentiate the elements of the cell by clustering. As for feature extraction we need to make clusters based on the elements in the cell so that we use the color based segmentation which gives us advantages.

1.4 Scope of Research

Our system detects the cancerous cell by testing a biopsy image. This system includes many procedures like color based segmentation, clustering and feature extraction which provides a vast field for working. With the help of this technique work can be done on **pattern recognition**, **Face detection or character identification**, Visual Inspection and Recognition also can be done. With the process of some normal work like feature extraction many information from a single image can be extracted, by using connected component counting any one can count the number of specified components in an image. In words through this system each of the technique in this work has many working field. In the field of **medical science** it has many scope of work and the topics of cancer detection is also an evolving and promising section of future works in image processing.

1.5 Overview

Chapter 2: Some terminologies and existing techniques have been discussed in this chapter.

Chapter 3: This chapter deals with the Proposed System and its implementation techniques along with its required methods and tools.

Chapter 4: In this chapter we have placed the Implementation and Testing of our proposed method.

Finally, in Chapter 5: We have concluded our work with some of its future development planning.

CHAPTER 2: BACKGROUND STUDY

2.1 About Image Processing

Image processing is a way or technique for analyzing an image and manipulating it. An image is a source of information. The technique of image processing is for finding out the information's and other necessaries by analyzing and manipulating the image. Actually the image processing focuses on two main tasks:

- Improvement of pictorial element for human interpretation
- Processing of image data for storage, transmission and representation for autonomous machine perception

So, actually image processing is a way for getting necessary information from an image or a way of making an image useful for some processing or work. Photographers perform certain image processing for making them attractive and visually outstanding.

Image processing actually predates modern computer technology. The first application was in newspaper industry, pictures were first sent by submarine cable between LONDON and NEW YORK. Introduction of the Bart lane cable picture transmission system in the early 1920s reduced the time required to transport a picture across the Atlantic from more than a week to less than three hours. An early motivation for the development of image processing techniques came from space program; in 1964 NSSA's jet **Propulsion Laboratory** used computers to correct distortions in the images of the lunar surface obtained by the Ranger 7 probe. Now more than three decades later image processing finds applications in areas as diverse as:

- Image enhancement and restoration
- Artistic effect
- Medical Visualization
- Industrial inspection
- Law enforcement
- Human computer interfaces

Our work is on the arena of medical visualization.

Image processing technique is a vast field and it involves lots of processes. The basic processing technique is similar for all the cases. The basic processing involves steps Image acquisition, enhancement, image restoration, Morphological processing, Segmentation, Recognition and Representation. The graphical representation:

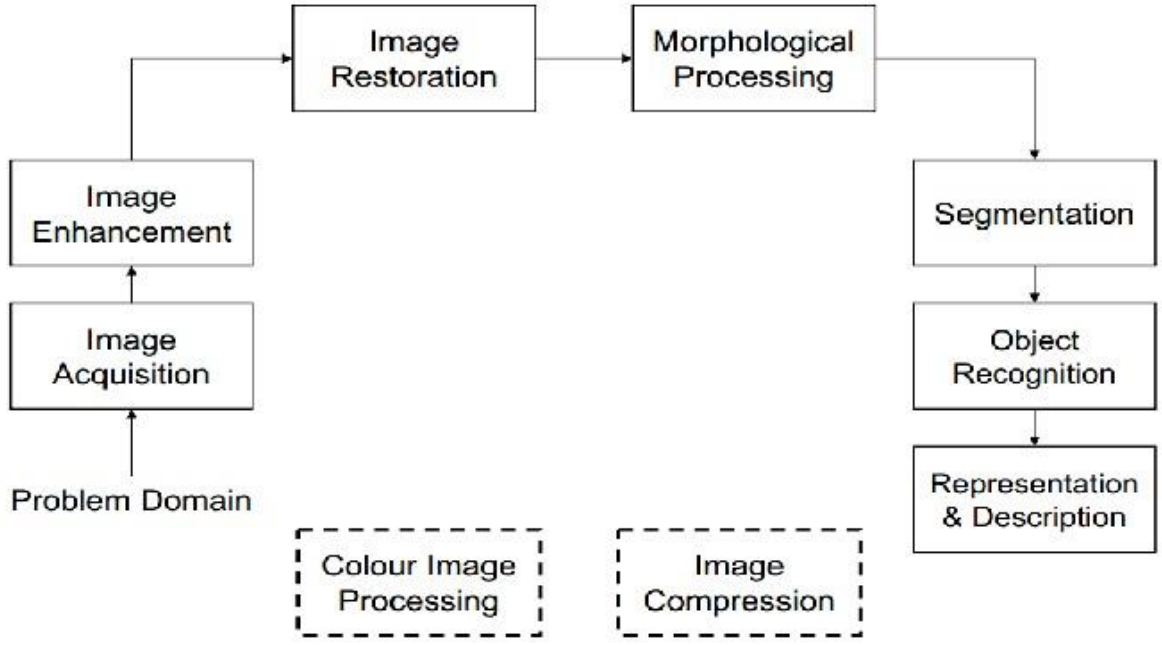

Fig 2.1: Basic Image Processing Steps

Image processing involves many processes like as Noise removal, Conversion, Segmentation, Morphological processing etc. We will discuss some of them whose are used in our work.

Type of Images

In this section the discussion will be the categories of medical image, how can we get these types of image and their use.

Medical images can be categorized according the way of getting them. Images found by ionizing radiation and non-ionizing radiation etc. The following figure shows us some categories of medical image:

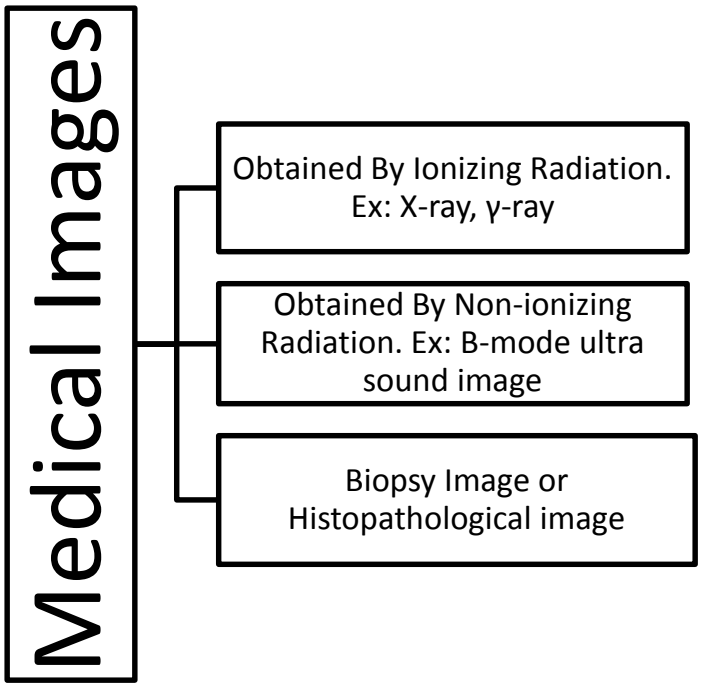

Fig 2.2: Some Types of Medical Images

The type Histopathological Image will be discussed later briefly in another section. Now other types are explained here.

Images Obtained By Ionizing Radiation

Medical Imaging technique produces images in accordance with arising signal or signals from the body of the patient. One of the several techniques is Ionizing radiation. Ionizing radiation in medical imaging comprises x-rays and γ-rays, both of which need to be used prudently to avoid causing serious damage to the body and to its geneticmaterial.

1. *X-ray Imaging*:X-ray imaging has been used in clinical diagnosis almost from the time of Roentgen's discovery of x-rays. X-rays are generated in an x-ray tube, which consists of an evacuated tube with a cathode and an anode. Images of the human body can be acquired or displayed in three main orientations as in following figure [11]:

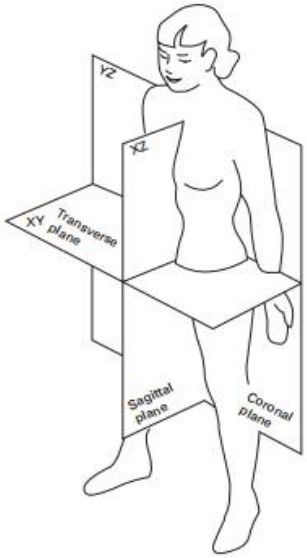

Fig 2.3: The three main imaging orientations for the human body: the coronal, sagittal and transverseplanes [11].

This is the orientation displayed in the common posterior–anterior (PA) chest radiograph, where the x-rays enter from the patient's back (posterior) and are collected by a film placed at his front (anterior) [1]. The acquired image is a superposition of many coronal images at different depths within the body. Thesagittalplane is a side view, dividing the body into right and left, and thetransverseplane, sometimes referred to as theaxialor trans-axial plane, is a plane perpendicular to the long axis of the body, dividing it into top and bottom planes.A typical normal posterior–anterior (PA) chest radiograph is shown in Figure:

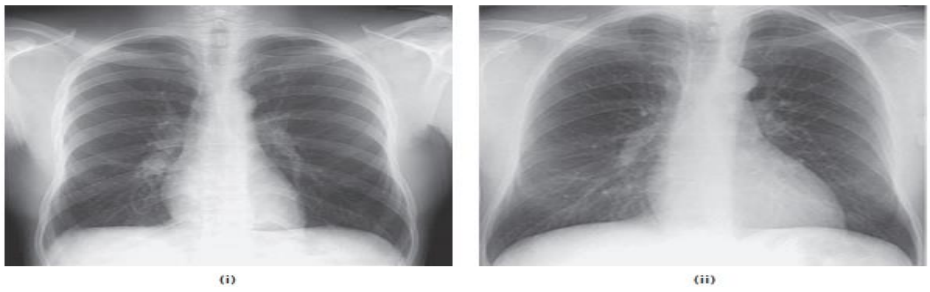

Fig 2.4: Posterior–anterior chest radiographs of (i) a normal patient and (ii) a patient suffering from tuberculosis.

CT scanner developed using the x-ray technology.CT imaging is the primary digital technique for imaging the chest, lungs, abdomen and bones due to its

ability to combine fast data acquisition and high resolution, and is ideally suited to three-dimensional reconstruction. It is particularly useful in the detection of pulmonary (i.e. lung) disease, because the lungs are difficult to image using ultrasound and MRI [11].

2. *γ-ray Imaging:* γ – imaging technology is a Nano-imaging technology.Nuclear medicine (NM) imaging uses theγ-rays emitted from radioactive isotopesattached to pharmaceutical tracers that are specific to certain physiological, metabolicand pathological activities, e.g. cerebral perfusion, myocardial perfusion, cancer.Theseradio-labeled pharmaceutical tracers are ingested or injected into the bodywhere they arecirculated and/or metabolized. Theγ-rays which they emit during radioactive decay passout of the body and are collected by detectors (gamma cameras) placed around thepatient; these measure the distribution of the tracer within the body, and produce imageswhich show the functionalor metabolic activity in the relevant organs [11]. The following figure shows how gamma camera work:

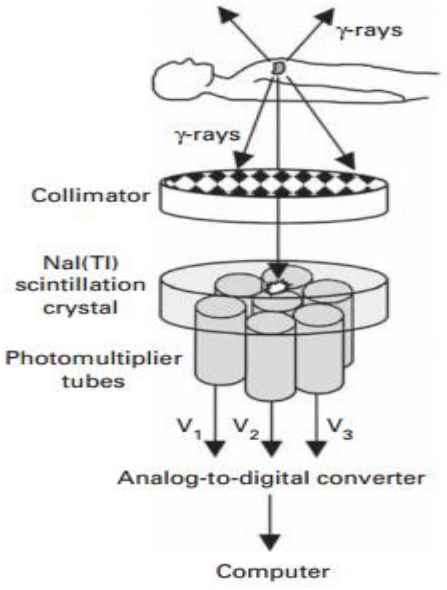

Fig 2.5: Schematic diagram for obtaining a planar nuclear medicine image, using a gamma camera [11].

Gamma imaging includes two types of imaging technique. They are: SPECT imaging and PET imaging.

SPECT imaging: single-photon emission computed tomography, SPECT, a rotating gamma camera, with one, two or three detector heads, rotates around and as closely as possible to the patient because spatial resolution decreases with the distance from the collimator. Sensitivity increases and acquisition time decreases with more detector heads [12].

The main advantage of SPECT over planar imaging is the absence of super positioning of overlying and underlying signals. A time-sequence of sequential SPECT images of the heart can easily be viewed as an animated sequence. SPECT studies of the brain are used to diagnose a large range of diseases that cause altered blood perfusion. SPECT scans can be used to measure cardiac wall thickness. Pseudo color is often added to the images to increase clarity. Some figure of them:

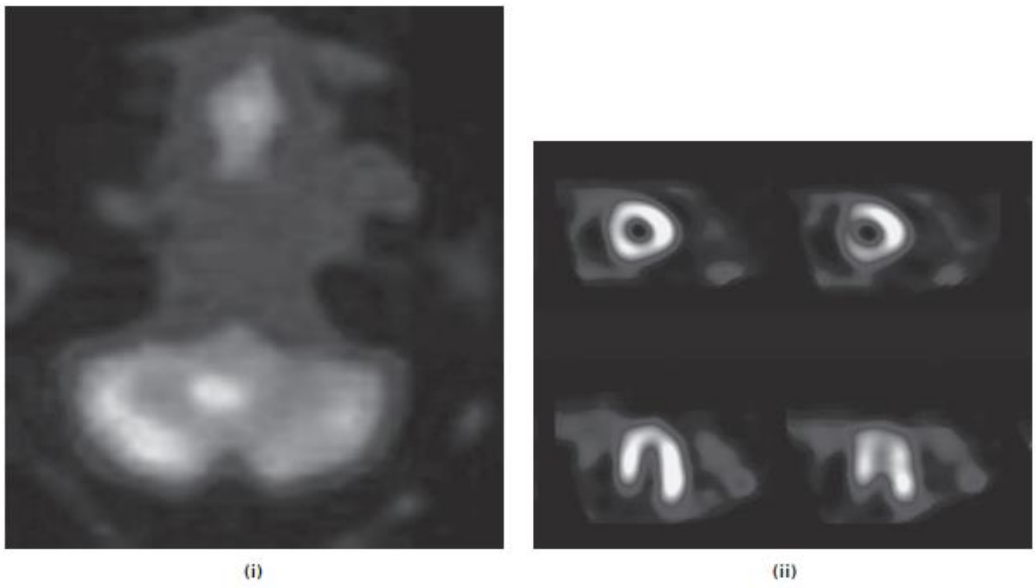

Fig 2.6: SPECT images showing (i) a brain tumor (in white), using 99mTc-GH , and (ii) thinning of the cardiac wall (reduced intensity), using 99m Tc-sestamibi [11].

PET imaging: Positron emission tomography, PET, is the most recent nuclear medicine imaging technique: in common with the others, it measures physiological function (e.g. perfusion, metabolism), rather than gross anatomy. A small, positron-emitting radioisotope with a short half-life (such as carbon-11,11C (about 20min), nitrogen-13, 13 N (about 10min), oxygen-15, 15 O (about 2min), and fluorine-18,18F (about 110min)) is incorporated into a metabolically active molecule (such as glucose, water or ammonia), and injected into the patient. Such labeled compounds are known asradiot racers**.**

It will more understandable if we look out the following figure:

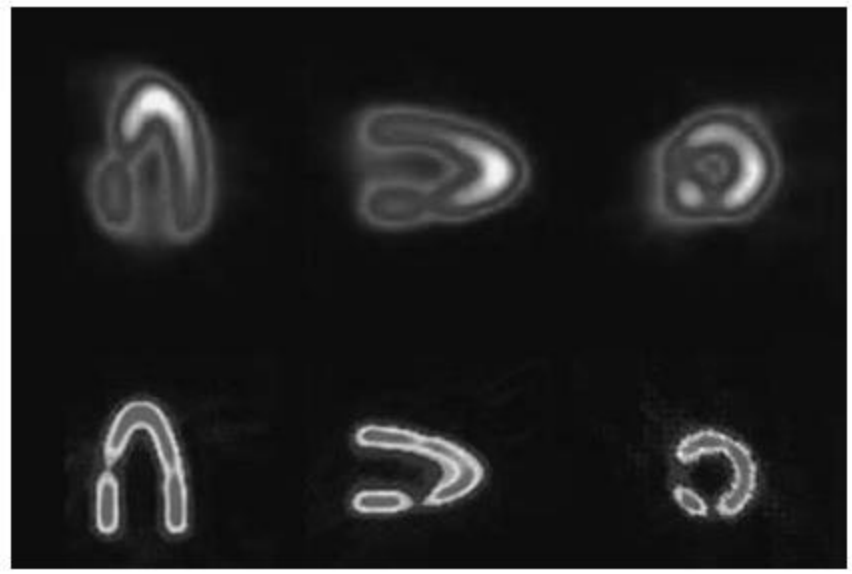

Fig 2.7: A realistic heart phantom imaged along three axes by SPECT with 99m Tc (top row) and PET with 18F-fluorodeoxyglucose (bottom row). See also color plate.

To facilitate the process of correlating structural and functional information, scanners that combine x-ray CT and radionuclide imaging, either SPECT or PET, have been developed. These dual-modality systems use separate detectors for x-ray and radio-nuclide imaging, with the detectors integrated on a common gantry. Because the two scans can be performed in immediate sequence during the same session, with the patient not changing position between the two types of scans, the two sets of images are more precisely registered.

Problems of Ionizing Radiation: There are some problems with images like x-ray and γ-ray images. X-ray and γ-ray photons have sufficient energy to ionize atoms and molecules within the body, causing serious and lasting biological damage. The absorbed dose, D, is equal to the radiation energy absorbed per unit mass of body; it is measured in units of grays (Gy) , where 1 gray is the dose when 1 joule of energy is absorbed per kilogram of irradiated material. However, the absorbed dose gives little indication of the risk to the patient; the biological damage caused also depends on the type of radiation.

Images Obtained By Non-Ionizing Radiation

Comparing to ionized image the non-ionized image is non-invasive. It does not create threat for the patient body. This is why now days this modality is largely used. There are several imaging in this modality like as: Ultra-Sound imaging and MRI (Magnetic resonance imaging) etc.

Ultrasound Imaging

Ultrasonic imaging uses high-frequency (~1–10MHz) sound waves and their echoes to produce images that can demonstrate organ movement in real time [2].Not like the X-ray andγ-ray the 'Ultrasound' technique is non-ionizing. It is considered safe at the intensities used in clinical imaging systems. Ultrasound images are constructed by calculating the time taken for ultrasound pulses to travel into the body and return, after reflection off a tissue surface.Three-dimensional ultrasound images can be obtained by adding additional rows of crystal elements to permit sweeping in a direction perpendicular to the plane of the B-mode scan.

Doppler imaging is included with it. By using the Doppler imaging we can measure the blood velocity. The Doppler Effect is familiar in the form of the increased frequency of a moving sound source, such as a train whistle or police siren, as it approaches, and the reduced frequency, as it passes by.

In the following figure we are seeing some image found by the Ultrasound technique:

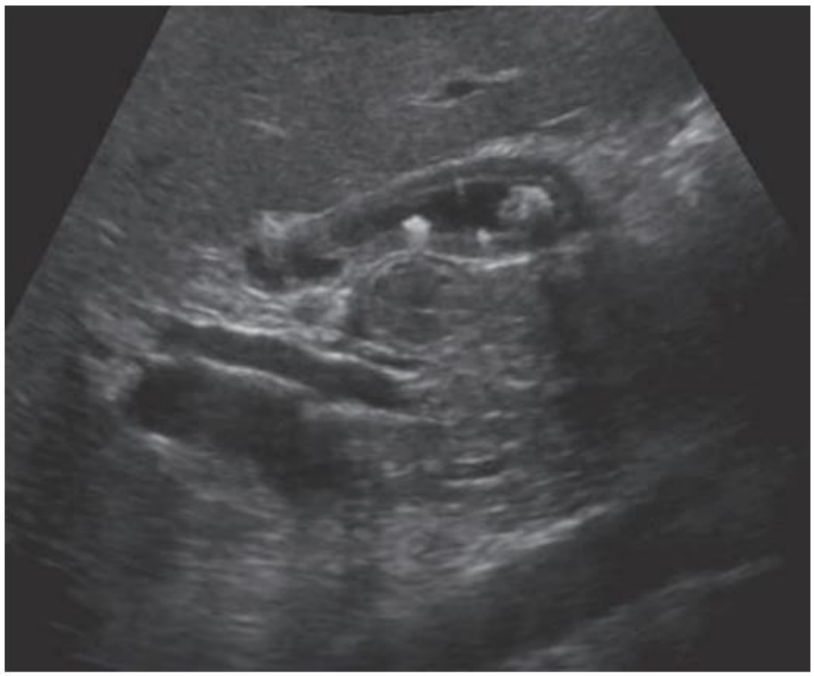

Fig 2.8: A B-mode image of a gall bladder, showing the presence of polyps [11].

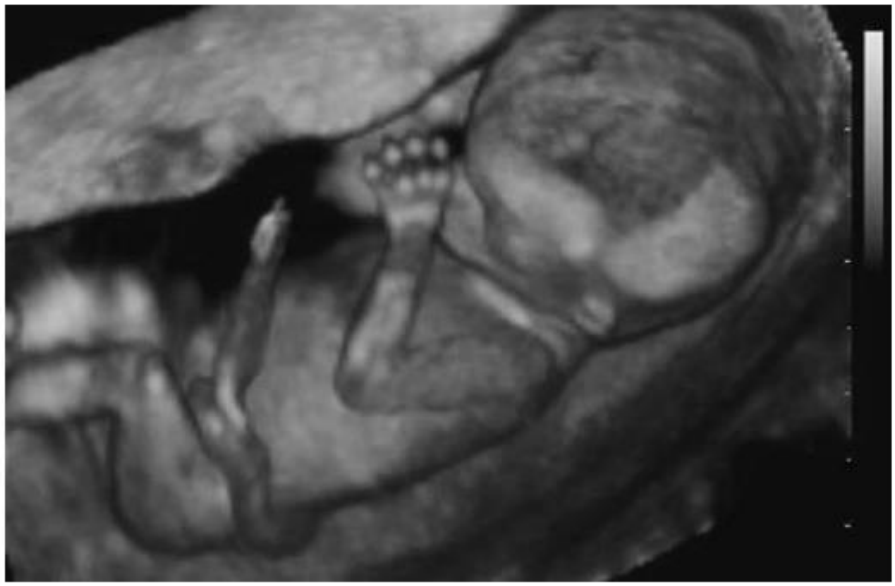

Fig 2.9: Rendered three-dimensional image of a 12-week old fetus [11].

There are a wide range of applications of ultrasound imaging as a result of its non-invasive, non-ionizing nature, and its ability to form real-time axial and three-dimensional images. The tissues of interest need to reflect sufficient ultrasound energy;

this limits the method to soft tissues, fluids and small calcifications preferably close to the surface of the body and unobstructed by bony structures [2].

Ultrasound is most commonly employed in examinations of the abdomen and pelvis. In obstetrics, fetal head size and fetal length are used as measures of fetal maturity and health, while spinal morphology can be used to detect the presence of abnormalities such as spina bifida. Doppler imaging can be used to measure fetal blood velocity and cardiac function [2].

MRI (Magnetic Resonance Imaging)

Magnetic resonance imaging (MRI) is a non-ionizing technique that uses radiofrequency (200MHz–2 GHz) electromagnetic radiation and large magnetic fields (around 1–2tesla (T), compared with the Earth's magnetic field of about 0.5×10 –4 T) [2]. The large magnetic fields are produced by superconducting magnets, in which current is passed through coils of superconducting wire whose electrical resistance is virtually zero [2].

The following image shows an image found by MRI technology:

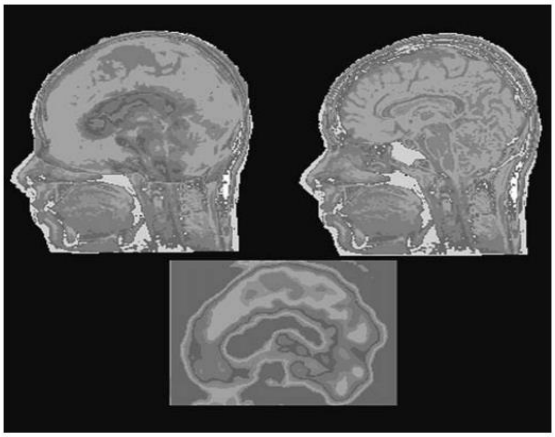

Fig 2.10: MRI (upper right) and SPECT (lower center) head sagittal slices of the same patient and the co-registered (MRI + SPECT) image (upper left). The lesion on the top of the skull is more prominent in the composite image, although it can be visualized in both modalities. See also color plate.

Histopathological Image

Histopathological images are one of the most important image types for medical analysis. It is mostly used in clinical diagnosis because it was fully safe and not causes any side effect for the body.
In later section (1.3) we will briefly describe about it.

2.2 Basic Image Conversion

Image Conversion is one of the most important part of basic image processing and a part of image analysis for diagnosis. As discussed in earlier section we already introduced with many types of image. Here we will now get introduced some new types which are used for the noise removal, analysis and other important purposes.

Some image for such use is **Binary Image** and **Gray Scale Image** etc. In computer systems image is a numerical status of a scene. An image is represented as a two-dimensional continuous light-intensity function, denoted by $i(x,y)$, where x and y represent spatial coordinates and the value or amplitude of i at any point (x,y) gives the intensity (brightness) of the image at that point.

In case of **Binary image** that has only two possible values for each pixel Black and White. The numerical value for black is 0 and for white is 1. On the other hand, in case of Gray scale image it also presented by black and white but it has some intensity level among black and white which is the advantage of Gray scale over binary image. Consider the following figure:

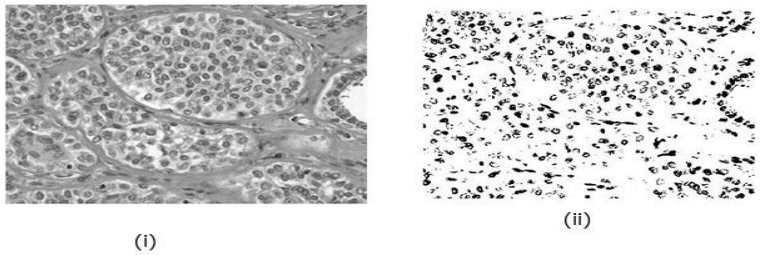

Fig 2.11: (i) gray image, (ii) binary image

Now the question is why conversion?

We need conversion of image for many reasons like when we are going to analyze an image for any purpose there we need to remove noise or cluster them or to take them in different color space. In accordance with the necessity we can convert any type of image to binary or gray or to other color space image.

Example: Converting a gray image to binary scale for removing noise. In the figure 1.12 we see image (i) is gray and image (ii) is binary. Only by seeing we can say that binary image is seems to be clearer than the gray, but in case of information gray is better than binary in some cases.

There also many conversion among different types of image are exists but giving their description in beyond this chapter. As we only discuss some basic conversion and their necessity.

2.2.1 Histopathological Image

Now in modern Medical Science the mostly used image types for Computer Assisted Diagnosis and Analysis Is Histopathological Image. This is also called clinical images. Actually histopathological imaging is come from the field histopathology. Actually Histopathology is the study of the microscopic structure of the diseased tissue. Histopathology refers to the microscopic examination of tissue in order to study the manifestations of disease [Wikipedia].

Histological analysis is performed by examining a thin slice (section) of tissue under a light (optical) or electron microscope (Jungueira&Carneiro, 2005, Kiernan, 2008, Ross et al., 2002, Mills, 2006). There are several techniques to produce histological images based on the analysis that will be conducted to detect diseased tissue. The way of getting histological image can be divided into two parts: First is tissue preparation and the second is image production.

Tissue preparation consist of fixation, dehydration, clearing, Infiltration, Embedding, Sectioning, staining and Slides [3]. The image production consists of getting the image of the tissue under the microscope. The following figure shows the full working map:

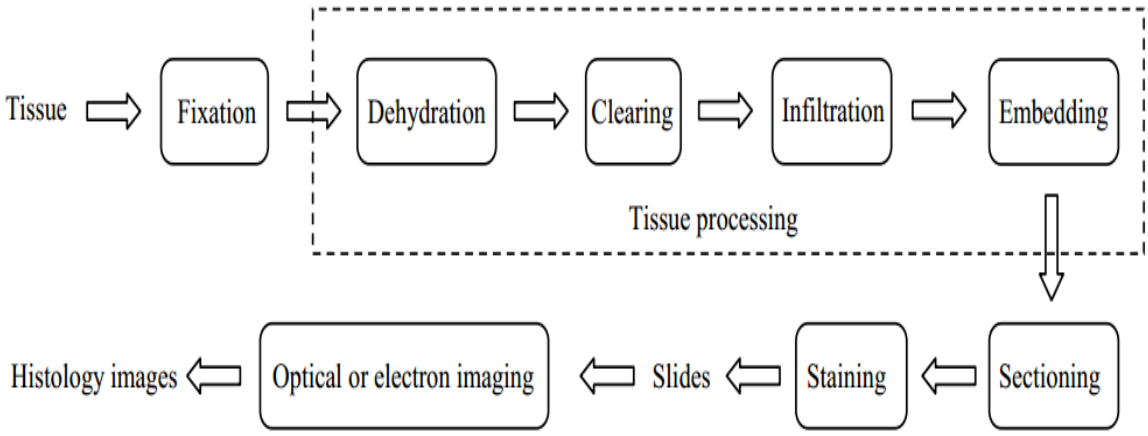

Fig 2.12: Getting histology images

The way of getting histology image is a linear as we see in the above figure describing the procedure of imaging.

The following figure shows histological image:

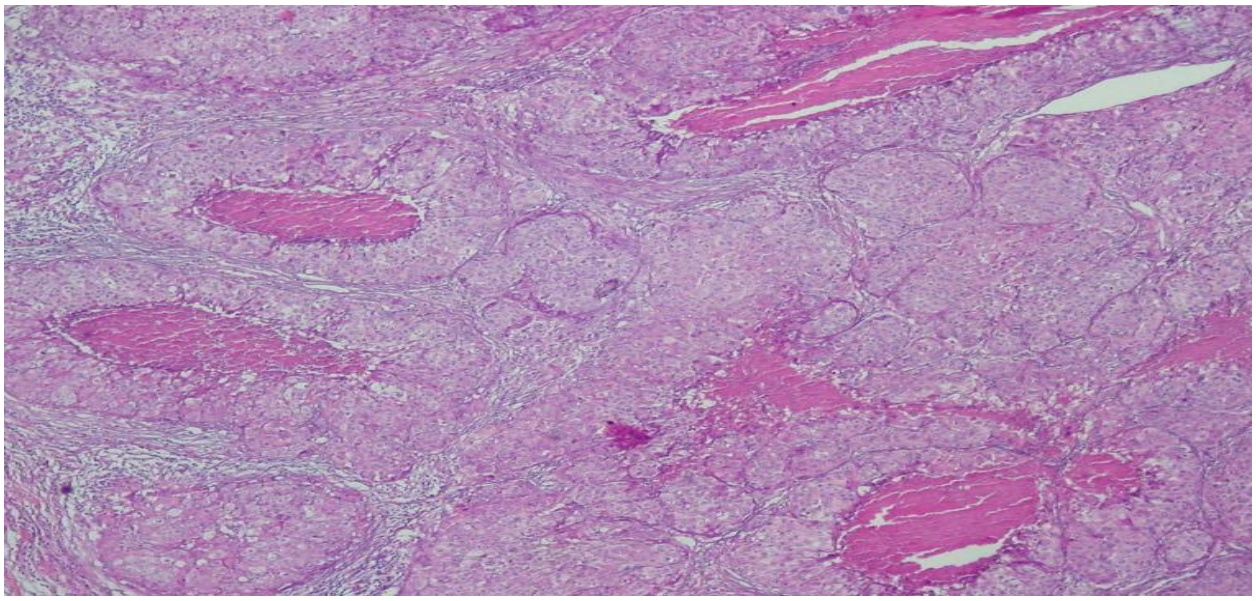

Fig 2.13: Histopathological image of cancer cell

Another Important Image is Biopsy Image. Biopsy means a sample of tissue from a human body. And the images that was founded by capturing after putting the tissue under a microscope.

The **Features of histopathological images** are very much useful for diagnosis and medical analysis. As this type of image is taken under a microscope so that each and every elements in the tissue come clear in actual color and size, which make it easier to analyze the tissue and find out its properties and necessary information.

Histopathological image is more significant than other images in computer based diagnosis. This type of images was used universally as like in cancer detection as Breast Cancer, Lung cancer and Cervical Cancer. And the expectation fulfilled by using this. Now most of the works in computer based diagnosis in this modern era are carried out over this type of image.

2.3 Machine Learning, Neural Network and the Brain Works

What is Machine Learning?

In general, machine learning involves adaptive mechanisms that enable computers to learn from experience, learn by example and learn by analogy. Learning capabilities can improve the performance of an intelligent system over time. Machine learning mechanisms form the basis for adaptive systems. The most popular approaches to machine learning are artificial neural networks andgenetic algorithms [15]. This chapter is dedicated to neural network.

2.3.1 Neural Network

A neural network can be defined as a model of reasoning based on the human brain [15]. We can define a neural network also as, a neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experimental knowledge and making it available for use [14].

It resembles the brain in two respects:

1. Knowledge is acquired by the network from its environment through a learning process,

2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.

A question can be raised that what are the benefits of using neural network? The answer is vast. But in short lists of some benefits of neural network are shown below:

1. Nonlinearity
2. Input-Output Mapping
3. Adaptivity
4. Evidential Response
5. Fault Tolerance
6. VLSI Implementation

Human Brain or Biological Brain Model

The human brain consists of a densely interconnected set of nerve cells, or basic information-processing units, called neurons. A neuron consists of a cell body, soma, a number of fibers called dendrites, and a single long fiber called the axon. While dendrites branch into a network around the soma, the axon stretches out to the dendrites and somas of other neurons. The following figure shows it:

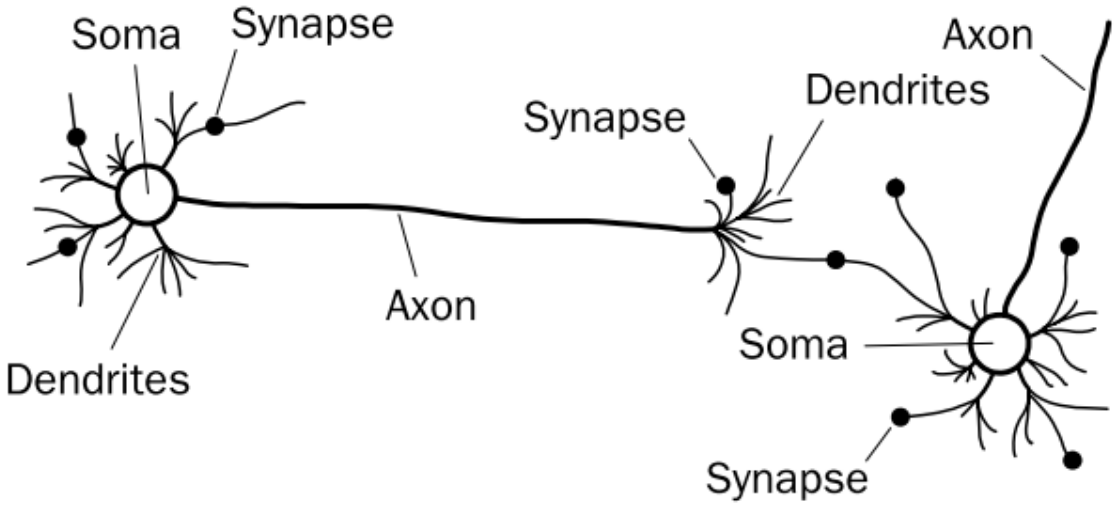

Fig 2.14: Human Brain (Real neuron Structure)

The human nervous system may be viewed as a three-stage system. Central to the system is brain, represented by the nerve system, which continually receives information, perceives it, and makes appropriate decisions [14].

2.3.2 Typical Artificial neural net or How Artificial neural net model the brain

An artificial neural network consists of a number of very simple and highlyInter-connected processors, also called neurons, which are analogous to the biological neurons in the brain.The architecture in steps [14]:

1. A set of synapses or connecting links, each of which is characterized by a weight or strength of its own.
2. An adder for summing the input signals.
3. An activation function for limiting the amplitude of the output of a neuron.

The following figure shows an artificial neural network:

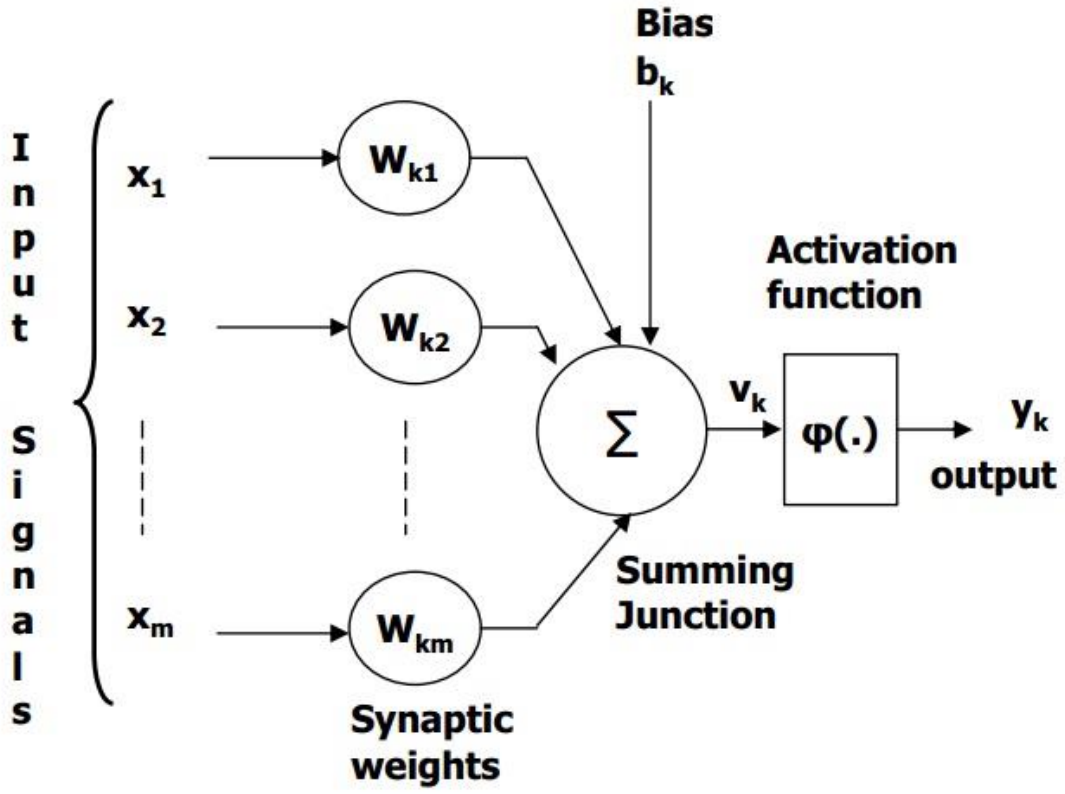

Fig 2.15: Basic Artificial Neural Network Architecture

Here x_1, x_2, $\ldots x_m$ are the input signals, w_{k1}, w_{k2}, . . ., $w_{km}$are the synaptic weights of neuron k, $u_k$is the linear combiner output due to the input signals, $b_k$is the bias, φ(.) is the activation function, and $y_k$is the output signal of the neuron [14].

How does Neural Network Learn?

The neurons are connected by links and each link has a numerical weight associated with it. Weights are the basic means of long-term memory inANNs.They express the strength, or in other words importance, of each neuron input. A neural network 'learns' through repeated adjustments of these weights [15].

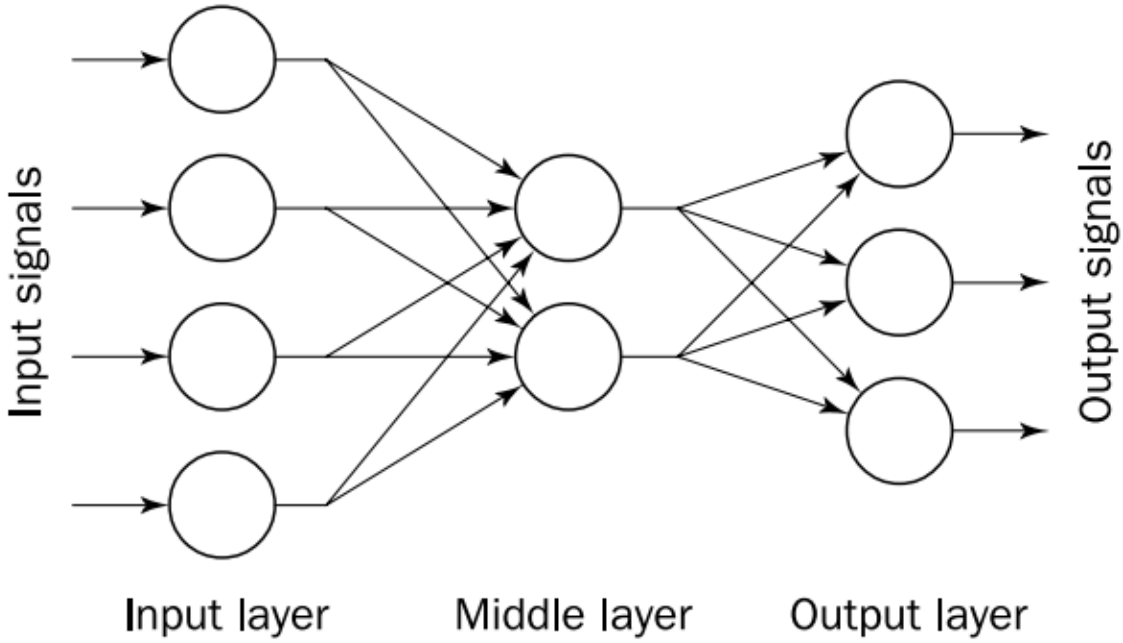

Fig 2.16: A typical neural network [15]

How It Adjust the Weight?

As shown in Figure 2.16, a typical neural network is made up of a hierarchy of layers, and the neurons in the networks are arranged along these layers. The neurons connected to the external environment form input and output layers. The weights are modified to bring the network input/output behavior into line with that of the environment. Each neuron is basic information processing unit.

To build an artificial neural network, we must decide first how many neurons are to be used and how the neurons are to be connected to form a network. Then we decide which learning algorithm to use. And finally we train the neural network, that is, we initialize the weights of the network and update the weights from a set of training examples.

Neuron as a Simple Computing Element in Artificial Intelligence

A neuron receives several signals from its input links, computes a new activationlevel and sends it as an output signal through the output links. The input signal can be raw data or outputs of other neurons. The output signal can be either a final solution to the problem or an input to other neurons. Figure 2.15 shows a basic neuron structure.

How Does Neural Network Determine Its Output?

In 1943, Warren McCulloch and Walter Pitts proposed a very simple idea that is still the basis for most artificial neural networks. The neuron computes the weighted sum of the input signals and compares the result with a threshold value. If the net input is less than the threshold, the neuron output is1. But if the net input is greater than or equal to the threshold, the neuron becomes activated and its output attains a value þ1 (McCulloch and Pitts, 1943) [15].

Neuron uses the activation function that is defined by the developer. There are several types of Activation function.

2.3.3 Types of Activation Function Used by the Neuron

Many activation functions have been tested, but only a few have found practical applications. Four common choices – the step, sign, linear and sigmoid functions – are illustrated in Figure 2.17.

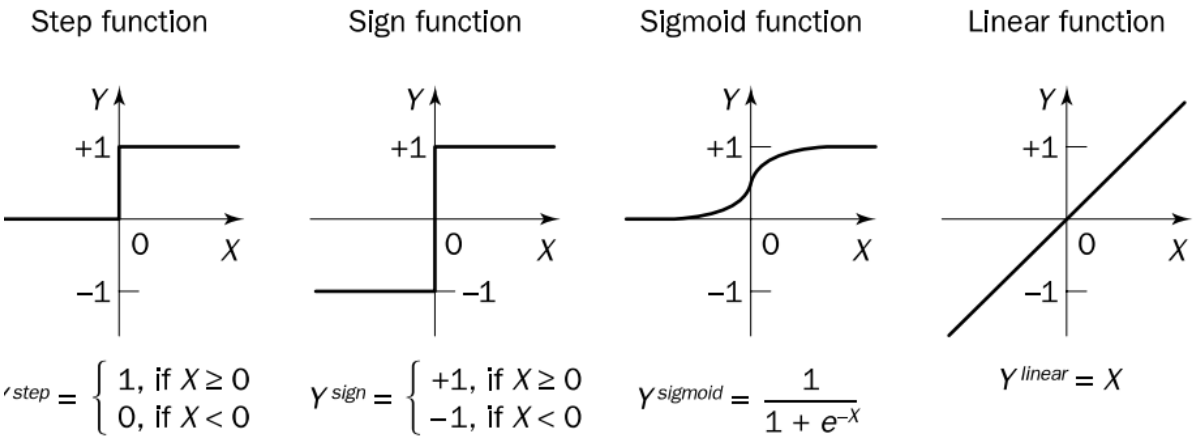

Fig 2.17: Activation Functions

Thestepandsignactivation functions, also called hard limit functions, are often used in decision-making neurons for classification and pattern recognition tasks [15].

2.3.4 Types of Neural Network

There are many types of neural network and most them will be presented in this section. But a question is arises, Can a single neuron learn a task?

The answer is, in 1958, Frank Rosenblatt introduced a training algorithm that provided the first procedure for training a simple ANN: a perceptron (Rosenblatt, 1958). The perceptron is the simplest form of a neural network. It consists of a single neuron withadjustablesynaptic weights and a hard limiter [15]. A single-layer two-input perceptron is shown in the following Figure.

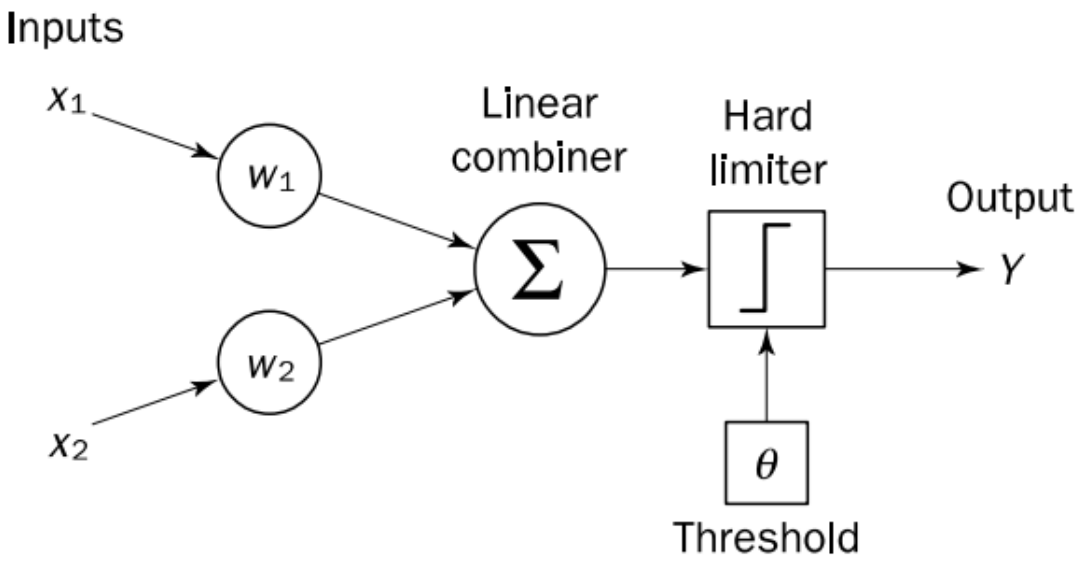

Fig 2.18: Single layer Two input perceptron [1]

The operation of Rosenblatt's perceptron is based on the McCulloch and Pitt's neuron model. The model consists of a linear combiner followed by a hard limiter. The weighted sum of the inputs is applied to the hard limiter, which produces anoutput equal toþ1 if its input is positive and1 if it is negative. Theaim of the perceptron is to classify inputs [15].

Single Layer Feed Forward Network

In a layered network the neurons are organized in the forms of layers. In the simplest form of a layered network, we have an input of source node that projects onto an output layer of neurons, but not vice versa [16]. The following figure shows a Single layer feed forward network.

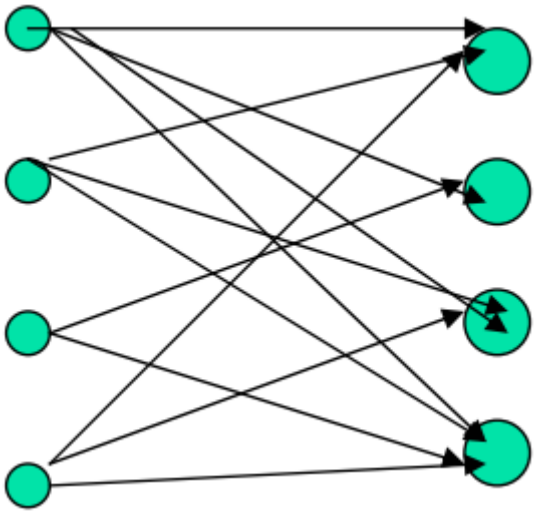

Fig 2.19: Single Layer Feed Forward Network

"Single layer" referring to the output layer of the computation nodes. We do not count the input layer because no computation is performed there.

Multi-Layer Feed Forward Network

This type of neural network distinguishes itself the presence of one or more hidden layers, whose computation nodes are correspondingly called hidden units. The function of the hidden neurons is to intervene between the external input and the network output in some useful manner. By adding one or more hidden layers, the network is enabling to extract higher order statistics [16].

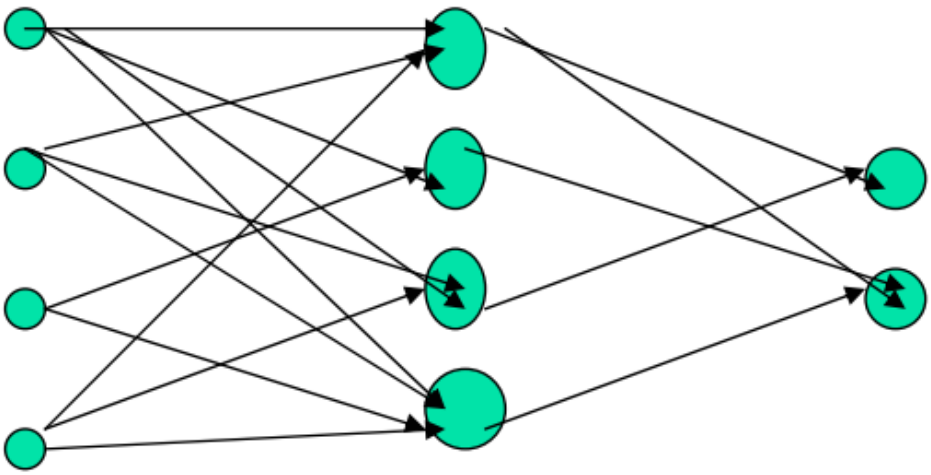

Fig 2.20:Multi-Layer Feed Forward Network

Recurrent Network

A recurrent neural network distinguishes itself from a feed forward neural network in that it has at least one feedback loop. For example, a recurrent network may consist of a single layer of neurons with each neuron feeding its output signal back to the inputs of all other neurons. A recurrent neural net

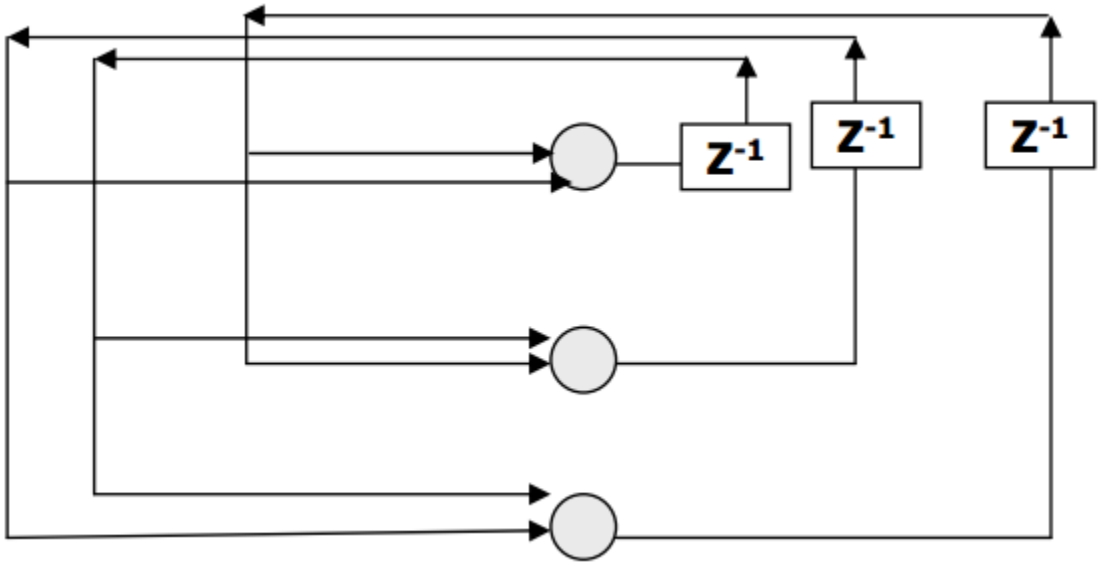

Fig 2.21: Recurrent Neural Net

In the recurrent neural network it may contain self-feedback loops. Self-feedback loops means where the output of the neuron is fed back to its own input. Another thing is that the recurrent also may have hidden neurons that mean it can be multilayer [17].

2.3.5 Learning System

What is Learning System

Learning is a process by which the free parameters of a neural network are adapted through a process of stimulation by the environment in which the network is embedded. The type of learning is determined by the manner in which the parameter changes take place.

There are different Types of learning algorithm. They are discussed in the following discussion.

Error-Correction Learning

We can define error-correction learning through the following equation:

$$e_k(n) = d_k(n) - y_k(n)$$

Where, k is a neuron, $y_k$is the output signal of neuron k, $d_k$is the desired response of k, $e_k$is the error signal.

Hebbian Learning

Hebbian Learning introduced by neurophysiologist Hebb (1949). The algorithm has two basic steps:

1. If two neurons on either side of a synapse are activated simultaneously then the strength of that synapse is selectively increased.

2. If two neurons on either side of a synapse are activated asynchronously then the synapse is selectivelyweakened.

Competitive Learning

In Hebbian learning, several output neurons may be active simultaneously, but in competitive learning only a single output neuron is active at any one time. It is the feature that makes competitive learning highly suited to discover statistically salient features that may be used to classify a set of input pattern.

Learning Rule:

1. A set of neurons that are all the samewith randomly distributed synaptic weights,

2. A limit imposed on the strength of each neuron

3. A mechanism that permits the neurons to compete for the right to respond to a given subset of inputs.

Supervised Learning

In the supervised learning system there will be a guider or teacher to guide the system to learn from the environment. The figure shows a supervised learning system.

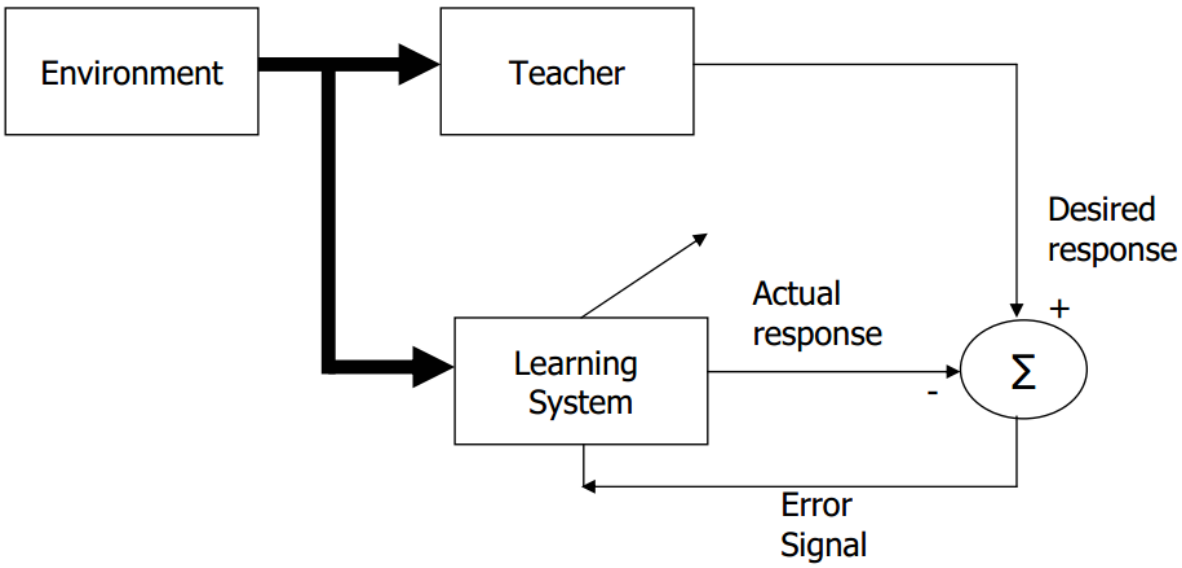

Fig 2.22: Supervised Learning System

Un-Supervised Learning

In the un-supervised learning system there will be no guider or teacher to guide the system to learn from the environment. The figure shows aun-supervised learning system.

Fig 2.23: Un-Supervised Learning System

2.4 Common Processing That Is Required To Determine Cancer Cell:

Cancer cell detection by microscopic image analysis is a work of Image processing. This is a sequential or step by step work. All of the existing techniques for this work are carried out by following some common hierarchical steps, but among the various cancer detection techniques differences is in the working procedure and algorithms in different steps. The steps includes Pre-processing, Analyzing the Image and Decision Making.

2.4.1 Pre Processing:

Pre- procession is the first step for this work. Pre-processing in image processing is an important part. After the image acquisition pre-processing is carried out. For this work pre-processing is required for the noise removal and cellular level feature extraction. In image processing there are many procedures for noise removal and enhanced the illuminati in the image for cellular level feature extraction.

There are different noise model and different procedures for removing the different noise. The procedure includes thresholding, histogram equalization and many filters for different noise model.

Thresholding: Thresholding is an image processing procedure that treats an image pixel as an object. In this technique pixel values above a threshold is set to a value and below the threshold set to another value. This is for enhancing the illuminati of the image.

Filters: There many different types of filters for different types of noise model. The noise model and filters are shown in following table:

Table 2.1 : Noise Model and Related Filters

Noise Model	Filters
Gaussian	Midpoint Filter
Uniform	Midpoint Filter
Impulse	Contra Harmonic Mean Filter
Periodic Noise	Band Reject Filter

Image Analysis

After pre-processing the next step is analyzing the image for finding out the characteristics of the image which is help full for determining whether the cell is cancerous or tumorous or normal. In image processing there are many techniques for analyzing the image, like as Morphology, Image Segmentation and image recognition and classification etc. Now some descriptions are given for these techniques.

Morphological Image Processing: Morphological Image processing techniques actually work with the shape and structures of the objects in an image.

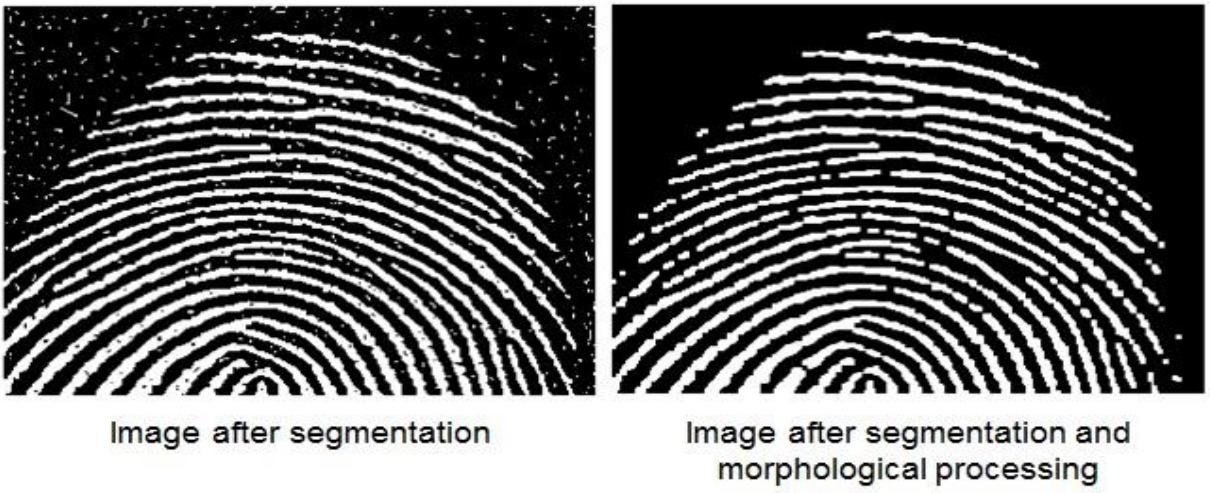

Fig 2.24: Example image of Morphological Processing of finger print [1]

Morphological image processing is only applicable for binary and gray scale images. There are two basic morphological processing:

(a) *Erosion:* Erosion of image *f* by structuring element *s* is given by $f \ominus s$
The structuring element s is positioned with its origin at *(x, y)* and the new pixel value is determined using the rule [1]:

$$g(x,y) = \begin{cases} 1 \text{ if } s \text{ fits } f \\ 0 \text{ otherwise} \end{cases}$$

Erosion is an important technique. It can split apart joined object, strip away extrusions and shrinks objects [12].

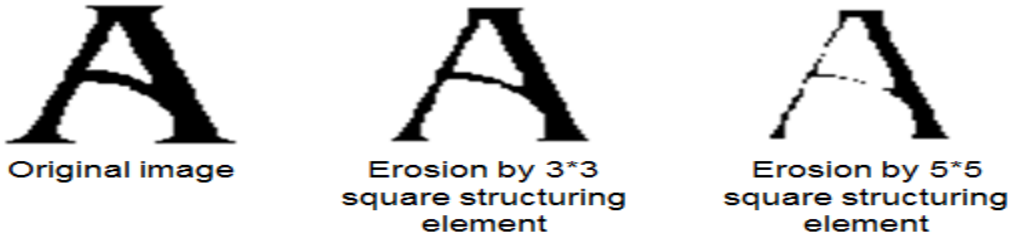

Fig 2.25: Example of Erosion [12]

(b) *Dilation:* Dilation of image *f* by structuring element *s* is given by *f s*The structuring element s is positioned with its origin at *(x, y)* and the new pixel value is determined using the rule:

$$g(x,y) = \begin{cases} 1 \text{ if } s \text{ hits } f \\ 0 \text{ otherwise} \end{cases}$$

This procedure is useful for repair breaks in image, repair intrusion and enlarges the objects [12].

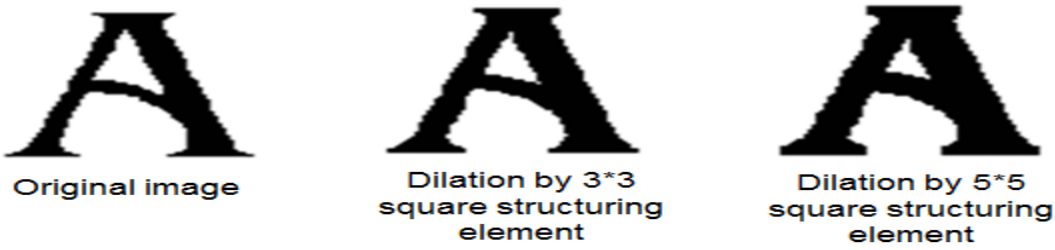

Fig 2.26: Example of Dilation [1]

There are also some other technique including above two as Opening, closing etc.

Morphological operators are non-linear, and common usages include filtering, edge detection, feature detection, counting objects in an image, image segmentation, noise reduction and finding the mid-line of an object [12].

Image Segmentation: Segmentation is the partitioning of an image into meaningful regions, most frequently to distinguish objects or regions of interest (“foreground”) from everything else (“back ground”) [12]. Or more easily it can be said that, Segmentation

attempts to partition the pixels of an image into groups that strongly correlate with the objects in an image [12].

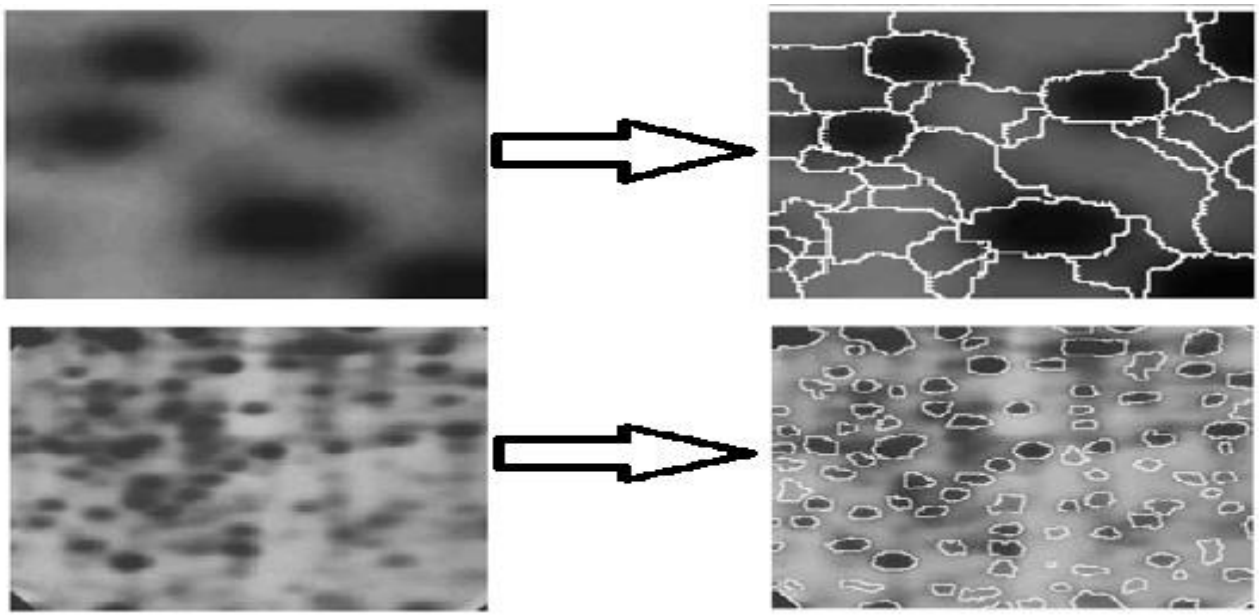

Fig 2.27: Example of image segmentation [11]

There are two major approaches for image segmentation they are Region based methods and Boundary based method.

(a) *Region based Method:* Region based method find connected region based on similarity of the pixels with in them. The objective of this procedure is to produce connected regions that are as large as possible and allowing some flexibility within each region [11].

(b) *Boundary based Method:* Boundary-based methods are based on finding pixel differences rather than pixel similarities. The goal is to determine a closed boundary such that an inside (the object or Foreground) and an outside (the background) can be defined [11].

Segmentation is necessary in cancer cell detection because, in order to develop robust interpretation systems, it is important to use as much relevant a priori information as possible during segmentation.

Feature Recognition and Classification: Feature recognition includes identifying the in the image. And Classification involves sorting objects in an image into separate classes, and is often the final step in a general image analysis process. Automated classification is fundamental to computer-assisted diagnosis in medical imaging and many other applications, such as robotic vision and speech recognition. Often the information available to make a decision is imprecise and frequently the decision procedures are statistical in nature. In such cases statistical approaches are used and the diagnostic accuracy of classification can be measured by receiver-operating characteristic (ROC) curves. However, if the fundamentals information is provided by the object structure then structural or syntactic methods are more appropriate. Recent methods, such as neural networks and genetic algorithms, borrow from both approaches [12].

A general diagnosis process can be figure out like the following:

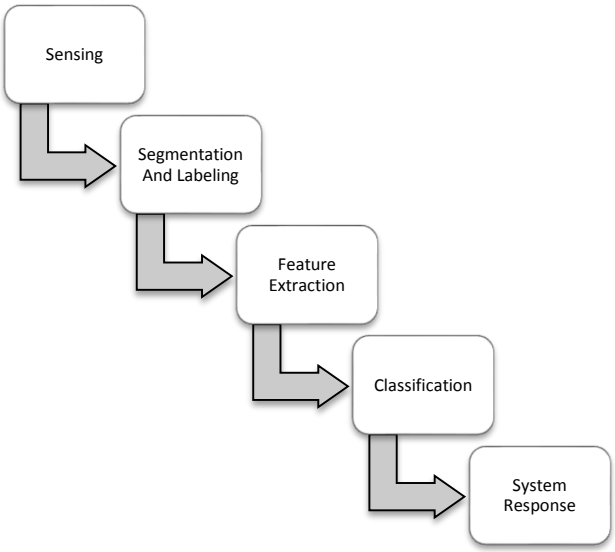

Fig 2.28: General Diagnosis Process [2]

2.4.2 Decision Making

Decision making is the critical step of cancer cell detection procedure. As for the cause that after feature extraction we will be able to know the characteristics of any cell and to develop a system for cancer cell detection we required the features of normal, tumor and cancer cell and teach the system about them because the cell image on which the system

going to investigate can be any type of image. So here is tuff job to make a system that will able to give accurate answer.

There already some procedure those are able to learn and take decision by comparing his learned information with reliable accuracy, Such as SVM (Support Vector Machine), Neural Network.

Neural Network

A neural network can be defined as a model of reasoning based on the human brain. Like human brain Neural network can learn by some mechanism and able to decide answer from the learning.

We can define neural network as, a neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experimental knowledge and making it available for use.

It resembles the brain in two respects:

1. Knowledge is acquired by the network from its environment through a learning process,
2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.

All about the neural network will be discussed in the chapter Neural Network.

2.5 Some Existing Techniques for Cancer Cell Detection

There are already many existing system for detecting cancer cell. Now some of them will be discussed in this section.

Automatic Cancer Diagnosis Based On Histopathological Images

Demir et al. shows a systematic survey for computational steps of the automated cancer detection. In his paper he talked about a technique. That is discussed below:

Automated cancer diagnosis at cellular level and tissue level are done based on-

1. Extracting information from Histopathological images of stained biopsies.

2. Examining the information by statistical analysis or machine algorithm.

This automated diagnosis consists of three computational steps:

1. *Preprocessing:* Eliminating noise and improve image quality is the task of this step. This also includes nucleus/cell segmentation incase of extracting cellular level information.

   Cellular-level feature extraction, noise reduction is followed by the segmentation process to determine the locations of the nuclei/cells1 in a tissue.
   The most trivial method for noise reduction is thresholding the pixels of an image which follows background correction and filtering.
   Filtering reduces the random noise and improves result for thresholding. The intensity histogram of an image is employed and the pixels under a threshold value are considered to be noise.

2. *Feature Extraction:* Features are extracted either at cellular level or at tissue level. In case of cell the morphological, textural, fractal, and/or intensity-based features are extracted. In case of tissue-level feature extraction quantifies the distribution of the cells across the tissue.

   In order to extract the features, there are two different types of information available in the image: (i) the intensity values of the pixels and (ii) their spatial interdependency.

3. *Diagnosis:* Distinguishing between benignity and malignancy, and classifying malignancy level are the task of this step.

Breast Cancer Detection and Classification of Histopathological Image

Singh et al. targeted job is detection of breast cancer and classifying the benign and malignant breast tumor. The technique used in their proposed system is:

Use d technique for carrying out the job to the goal:

1. Gray scale conversion
2. Contrast limited Adaptive histogram equalization

3. Adjust Image intensity
4. Adaptive thresholding based segmentation and morphological operation
5. Watershed segmentation and Morphological operation
6. Blob labeling
7. Feature extraction
8. Normalize feature vector
9. Design, training, validation and testing of feed forward back propagation neural network for breast cancer tumor classification
10. Cancer and biopsy image classification using feed forward back propagation neural network

Levenburg-Marquards algorithm is used for training the neural. Sigmoid and hyperbolic tangent functions are applied to the learning process. Feed forward back propagation neural network is used for classifying as type1, type2, type3. Input vector and corresponding target vector is used for training feed forward back propagation neural network.

Microscopic Image Segmentation and Recognition on Cancerous Cells

Swarnalathaet al. (2010) purpose of the work is to develop semi-automated system for lung cancer identification. According to their work the technique is:

They design a linear module consist of six module in which each module output is an input of next module. The modules are:

1. Digital Cytological Image: It is the System input.
2. Pre-processing: This is done by histogram approach. Improvement of image quality, finding representation of cell, enhance as for object are touching each other.
3. Segmentation: isolating objects. Watershed transformation method is used for segmentation.
4. Feature extraction: It is done by Morphometry.
5. Classification and Recognition: Back Propagation Neural Network (BPNN) is used for Classification.

6. Benign or Malignant Cell

They implement their work in MATLAB. A 256×165 image takes 3 minutes processing for answer.

Processing Technology in Microscopic Image of Cancerous Cells in Pleural Fluid Based on Fuzzy Edge Detection Method

Zhang et al. (2006) proposed an image processing technique for detecting cancerous cell in pleural fluid. Their technique:

Cancerous cell detection in pleural fluid depends on the shape on karyon.

Processed used for the work:

Original grayscale image ➔ OTSU threshold method ➔ Fuzzy edge detection ➔ Morphological analysis

Conventional edge detection is not seems to be suitable for this work so that fuzzy edge detection is used. Firstly images are enhanced in fuzzy characteristics domain. Inverse function of membership function is used to enhance image. Then extract edges.

After that need to calculate:

1. Area rate of karyon and Cytoplasm
2. Optic density of karyon

When above procedures are done further processing and experiment are carried out. The result of the experiment proves that this is feasible and effective cell detection method.

Microscopic Image Analysis and Recognition On Pathological Cells

Liu et al. proposed a work for segmenting cell and nucleus from a complex microscopic image, Automatic threshold segmentation is used which capable of segmenting image based on RGB and HIS color Space. Their technique in detail:

Therapy for a disease depends on the stage that is found through diagnosis. Throughout the paper cells on initial states called pathological cell.

They have used new approach for analyzing cancer cell based on segmentation and feature extraction using some technical algorithms.

Firstly, noises are reduced from image and make it adopt using threshold segmentation based on RGB and HIS ("H" for different color, "S" for Saturation and "I" for Intensity). After that using canny operator they detect pathological cell and nucleus from image background. Then extract feature using eight-chain code tracking technique.

After that they extract the features of normal cell for comparison. Then they compare using algorithm made for the comparison by them. For achieving higher accuracy color features are adopted for further classification.

Computer Assisted Diagnosis in Histopathology

Through the technique before the analysis each tissue is making prepared by some technical procedure that includes: Fixation, Infiltration, Dehydration, Embedding, Clearing, Sectioning and Staining.

State-Of-Art image segmentation (active contour) is used for feature extraction and pattern classification.

Histology Imaging Technologies: Varity of imaging technologies are used to take histological images with light or electron microscope. Some of them:

a. Light Microscopy: It is commonly used to magnify tissue structure and high resolution image. Its fundamental components include illumination system. It have four major type:
    i. Fluorescence Microscopy
    j. Confocal Microscopy
    k. Hyper spectral & Multispectral Microscopy
b. Electron Microscopy: For higher resolution of viewing smaller object requires more expensive electron microscopy. Some types:
    1. Transmission Electron Microscopy
    2. Scanning Electron Microscopy

Image Processing & Analysis: This system consists of conventional image processing including preprocessing, segmentation, feature extraction, feature dimension reduction, feature based classification and post processing.
Preprocessing is applied to reduce computational cost and obtain low resolution image.

Segmentation: Extracts object of interest from the background. These are focused for further cancer identification and classification.

Feature Extraction and Dimensionality Reduction For disease Identification:
In CAD system after segmentation features are extracted from region of interest to detect and grade potential cancer.

CIN Detection and Classification: Cervical carcinoma refers to the cancer forming in cervix tissues, which is always caused by human papillomavirus (HPV) infection.

In 1968, a standard nomenclature was introduced to classify the malignancy degree of cervical cancer, namely cervical intraepithelial neoplasia. CIN is divided into three grades.

One commonly used feature to determine the CIN grade is the nuclear-cytoplasm ratio: a larger ratio corresponds to a more severe CIN degree. Final diagnosis of CIN is conducted by histopathology analysis of cervix tissues. This analysis is based on the proportion of the epithelium with immature and undifferentiated cells.
In (Keenan et al., 2000), a machine vision system is introduced, which automatically grades the CIN degree given a histological image.
The method is as follows. After thresholding to locate all nuclei, Delaunay triangulation mesh is constructed using the centers of the nuclei. Morphological features including the mean triangle area, the mean triangle edge length, and the number of triangles in a unit area, are computed to determine the CIN grade. In total 18 features are derived and LDA is applied to select the most discriminating features for differentiating (1) normal and CIN 3, (2) koilocytosis and CIN 1, and (3) all CIN cases.

CHAPTER 3: PROPOSED SYSTEM

Introduction

This chapter is about the process in which the work of Cancer detection was done. The chapter also includes the stechniques and reasons of using these techniques. Through the chapter reader can know about noise removal, thresholding, feature extraction from an image and some technique of using data mining techniques.

3.1 Basic Steps of the Proposed System

Some basic steps through which the cancer detection work was done are discussed in the previous chapter (chapter 3). This work also following some of these basic steps and some new techniques with the basic steps. For decision making the data mining approach is used. The steps of the work are listed below:

- Transforming the image in different Color Space
- Color Based Segmentation and Clustering
- Segment The Blue Nuclie into Separate Image From Any Cluster
- Noise removal
- Feature Extraction
- Decision Making

Through these steps the process of Malignant Cancer Cell Detection By Microscopic image analysis was done. Now the following section will discuss about each of the steps.

3.2 Description of the Steps

The working procedure and discussion about them are added below sections.

3.2.1 Transforming the Image in Different Color Space

The first step of this work is converting the image in different color space. The conversion was done because as known to all that the different elements in the cell are off different color and color also varied with the density off the liquid element, as the target is clustering on the basis of the elements of the cell so the conversion was done.

The Matlab code for doing this color space conversion:

```
he = imread('Histo.jpg');
cform = makecform('srgb2lab');
lab_he = applycform(he,cform);
```

Here the first line i …g. In the second line the work of **makecform()** funct…ation structure **cform** that defines the color space co… form the transformation, pass the color transformation structure as an argument to the **applycform** function. The type "**srgb2lab"** Convert from the *sRGB* to the *L*a*b** color space.sRGB is a standard RGB (Red, Green and Blue) color space and L*a*b is a Lab color space which is a color-opponent space with dimension L for lightness and a and b for the color-opponent dimensions, based on nonlinearly compressed CIE XYZ color space coordinates.

The following figure shows an image in sRGB and an image in L*a*b color space:

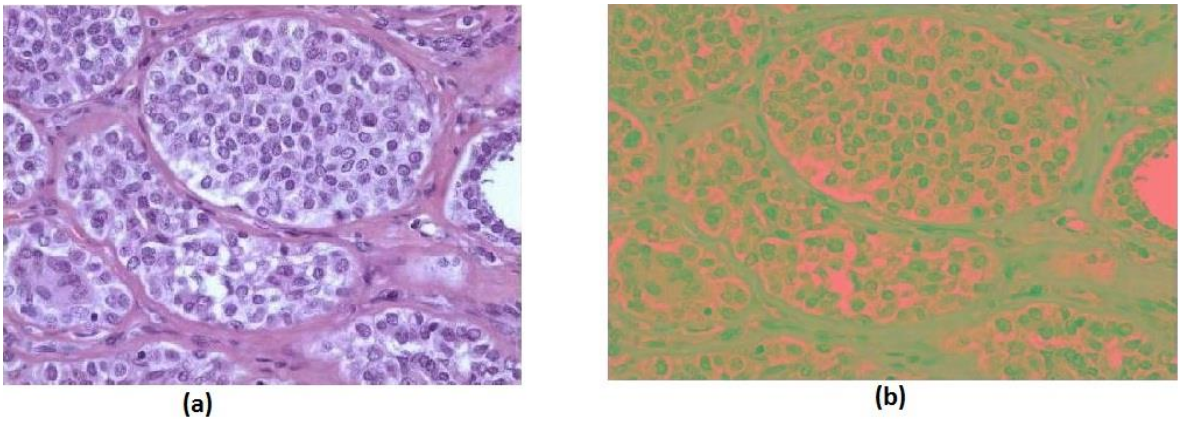

Fig 3.1: (a) Image in sRGB color space and (b) Image (a) is converte to L*a*b color space

3.2.2 Color Based Segmentation and Clustering

After converting the image in to L*a*b color space the next work is to segment and cluster the image into three different color on the basis of the color because colors are indicating different elements within the cell image.

The Matlab code:

```
ab = double(lab_he(:,:,2:3));
nrows = size(ab,1);
ncols = size(ab,2);
```

```
ab = reshape(ab,nrows*ncols,2);
nColors = 3;
[cluster_idx,cluster_center] =
kmeans(ab,nColors,'distance','sqEuclidean','Replicates',3);
pixel_labels = reshape(cluster_idx,nrows,ncols);
imshow(pixel_labels,[]), title('image labeled by cluster index');
```

In this code firstly the size that means the number of rows and columns are taken in variable ab. Then size of the rows and columns are declared. After that clustering is done using Kmeans clustering and repeat the clustering three times to avoid local minima. The kmeans clustering is used because, K-means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. K-means clustering requires that you specify the number of clusters to be partitioned and a distance metric to quantify how close two objects are to each other.

Using pixel_labels, the separation objects are done in the image. The last portion of the code shows the clusters of the image.

The following figure shows cluster of the image showed in Fig 6.1(a).

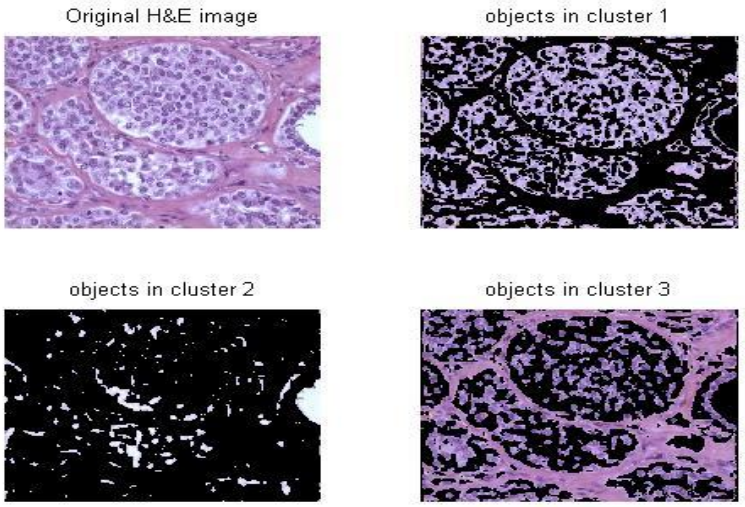

Fig 3.2: Image After clustering

3.2.3 Segment The Blue Nuclie into Separate Image From Any Cluster

Notice that there are dark and light blue objects in one of the clusters. To separate dark blue from light blue using the 'L*' layer in the L*a*b* color space. The cell nuclei are dark blue.

Recall that the 'L*' layer contains the brightness values of each color. Find the cluster that contains the blue objects. Extract the brightness values of the pixels in this cluster and threshold them using im2bw.

You must programmatically determine the index of the cluster containing the blue objects because kmeans will not return the same cluster_idx value every time. You can do this using the cluster_center value, which contains the mean 'a*' and 'b*' value for each cluster. The blue cluster has the smallest cluster_center value.

Matlab code:

```
 mean_cluster_value = mean(cluster_center,2);
[tmp, idx] = sort(mean_cluster_value);
blue_cluster_num = idx(1);


L = lab_he(:,:,1);
blue_idx = find(pixel_labels == blue_cluster_num);
L_blue = L(blue_idx);
is_light_blue = im2bw(L_blue,graythresh(L_blue));
```

Then the mask is used is_light_blue to label which pixels belong to the blue nuclei. Then display the blue nuclei in a separate image.

Matlab code:

```
nuclei_labels = repmat(uint8(0),[nrowsncols]);
nuclei_labels(blue_idx(is_light_blue==false)) = 1;
nuclei_labels = repmat(nuclei_labels,[1 1 3]);
blue_nuclei = he;
```

```
blue_nuclei(nuclei_labels ~= 1) = 0;
imshow(blue_nuclei), title('blue nuclei');
```

The following image shows the blue nuclie of the image that was read and showed in 3.1(a).

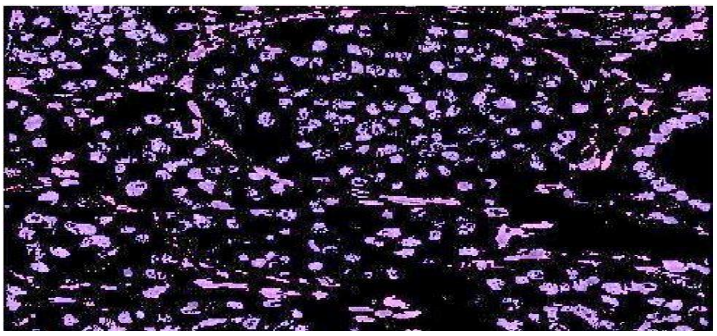

Fig 3.3: Blue Nuclie Image

3.2.4 Noise Removal

Noise removal is the most important part of the cancer detection process. Because if noise is not removed properly the feature extraction will contain noisy data and decision making will be insist by noise as for which we will not be able to get proper output.

There are different types of noise and different techniques for removing these noises. As we are working on biopsy image and most of the cases they are contain a minimum amount of noise so don't have to use any critical algorithm or any complex technique. As the image was converted in L*a*b color space, so noise will get different color and in blue nuclie, image it will approximately different in color. So that we just make our required color object position one and other positions zero by doing gray level thresholding. This process gives us noise free image.

Matlab code:

```
level = graythresh(I2);
bw = im2bw(I2,level);
```

After noise removal the blue nucli image will be look like:

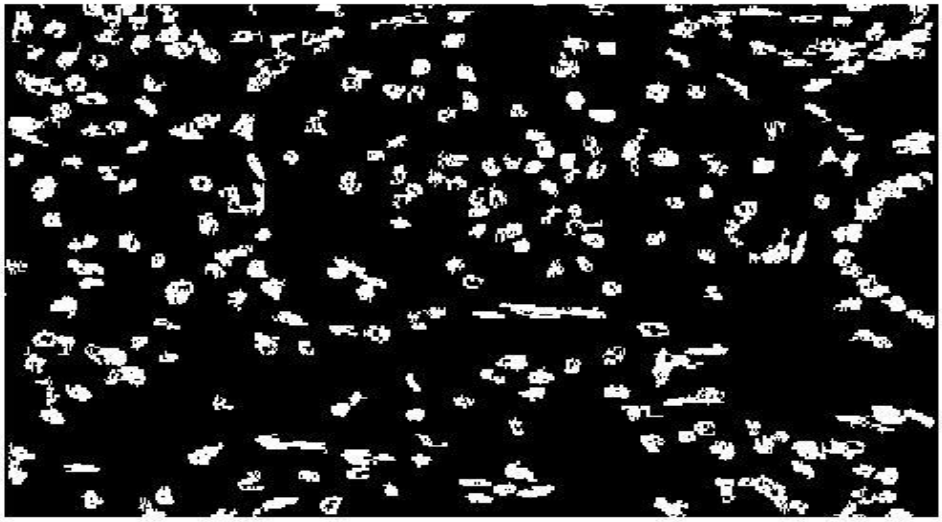

Fig 3.4: After noise removal the blue nuclie

3.2.5 Feature Extraction

Feature extraction is the most important part of the whole procedure. In this portion from the image different properties of the cell will be extracted. As like, we find image containing blue nuclei, so number of nucleus can be find by counting the number of connected components.

In this work of cancer detection some feature are extracted from image. They are described below.

First the numbers of connected components are counted for assuming the number of nucleus in that cell.

```
bw = bwareaopen(bw, 50);
figure, imshow(bw)
cc = bwconncomp(bw, 4)
cc.NumObjects
```

bwconncompreturns the connected components CC found in BW. The binary image BW can have any dimension. CC is a structure with four fields.**bwconncomp** uses a default connectivity of 8 for two dimensions, 26 for three dimensions, and **conndef(ndims(BW),'maximal')** for higher dimensions.

CC = bwconncomp(BW,conn) specifies the desired connectivity for the connected components. conn can have scalar values like 4 for four connected neighborhood, 8 for eight connected neighborhood etc.

Next work is finding the number of black pixel and white pixel in that image. The Matlab code:

```
xmax=800;
ymax=800;
 BW1=double(bw);%Here bw is a binary Image
White_pix=0;
Black_pix=0;

Floc=0;
for j=1:(xmax)
for i=1:(ymax)
if BW1(i,j)==1
White_pix=White_pix+1;
%if need to count black pixel
else
Black_pix=Black_pix+1;
end
end
end
```

This portion code find the number white pixel and black pixel by checking each coordinate value of the image.

Then we find the pixel ratio by dividing black pixel with white pixel. These all are work as parameter for our decision making in the next step.

3.2.6 Decision Making

In our procedure there are three classes, first is normal, second one is insitu and third one is invasive. By getting information from all data we made a range of value by data mining process, if the any image which will be checked have the values of the parameter within the range of first class then it is normal image, if not then we checked for second class if match their it will be insitu that means it's a tumor cell and going to grow as a cancer cell and if not in second class then we check for third class and if found match their than it will called as malignant cancer cell.

The data mining algorithm which we used are given below:

Input: D, a database of transactions;

Output: L, frequent matched item sets in D

Method:

(1) L1= find frequent 1-itemsets(D);

(2) for (k = 2; Lk 1 6 = φ; k++) f

(3) Ck= apriorigen(Lk 1);

(4) for each transaction t 2 D f // scan D for counts

(5) Ct = subset(Ck , t); // get the subsets of t that are candidates

(6) for each candidate c 2 Ct

(7) c.count++;

(8) }

(9) Lk = fc 2 Ckjc:count>= min supg

(10)}

(11) return L = [kLk ;

procedureapriori gen(Lk 1:frequent (k 1)-itemsets)

(1) for each itemset l1 2 Lk 1

(2) for each itemset l2 2 Lk 1

(3) if (l1[1] = l2[1]) ^ (l1[2] = l2[2]) ^ ::: ^ (l1[k 2] = l2[k 2]) ^ (l1[k 1] < l2[k 1]) then f

(4) c = l1 on l2; // join step: generate candidates

(5) if has infrequent subset(c, L

k 1) then

(6) delete c; // prune step: remove unfruitful candidate

(7) else add c to Ck ;

(8) }

(9) return Ck ;

procedure has infrequent subset(c: candidate k-itemset;

L

k 1: frequent (k 1)-itemsets); // use prior knowledge

(1) for each (k 1)-subset s of c

(2) if s 62 Lk 1 then

(3) return TRUE;

(4) return FALSE

procedure has infrequent subset(c: candidate k-itemset;Lk 1: frequent (k 1)-itemsets);

(1) for each (k 1)-subset s of c

(2) if s 62 Lk 1 then

(3) return TRUE;

(4) return FALSE;

This is a pseudo-code for the Apriori algorithm and its related procedures. Step 1 of Apriori finds the frequent 1-itemsets, L1. In steps 2 to 10, Lk 1 is used to generate candidates Ck in order to find Lk for k >=2. The apriori gen procedure generates the candidates and then uses the Apriori property to eliminate those having a subset that is not frequent (step 3). This procedure is described below. Once all of the candidates have been generated, the database is scanned (step 4). For each transaction, a subset function is used to find all subsets of the transaction that are candidates (step 5), and the count for each of these candidates is accumulated (steps 6 and 7). Finally, all of those candidates satisfying minimum support (step 9) form the set of frequent itemsets, L (step 11). A procedure can then be called to generate association rules from the frequent itemsets.The apriori gen procedure performs two kinds of actions, namely, join and prune, as described above. In the join component, Lk 1 is joined with Lk 1 to generate potential candidates (steps 1 to 4). The prune component (steps 5 to 7) employs the Apriori prop-erty to remove candidates that have a subset that is not frequent. The test for infrequent subsets is shown in procedure has infrequent subset.

Summary

In this chapter the discussion is carried out over the whole process of cancer detection by biopsy image analysis. In each step the brief discussion is given in where, which

procedure has been used and the reason for choosing that technique and also the explanation of the algorithms.

CHAPTER 4: IMPLEMENTATION AND TESTING

4.1 Implementation

As it is a semi-automated system first some portion of the work done using MATLAB. We develop function and work on the selected image. First color space is converted, then segmentation is done and then clustering is need to do. After that manually select the proper cluster and then highlight the blue nuclei in image. Next work is taking the image as an input on our software.

For implementing the system Java technology is used. At first for developing the user interface and making the user interface easily workable we use Nimbus look and feel.

After implementing the user interface look like:

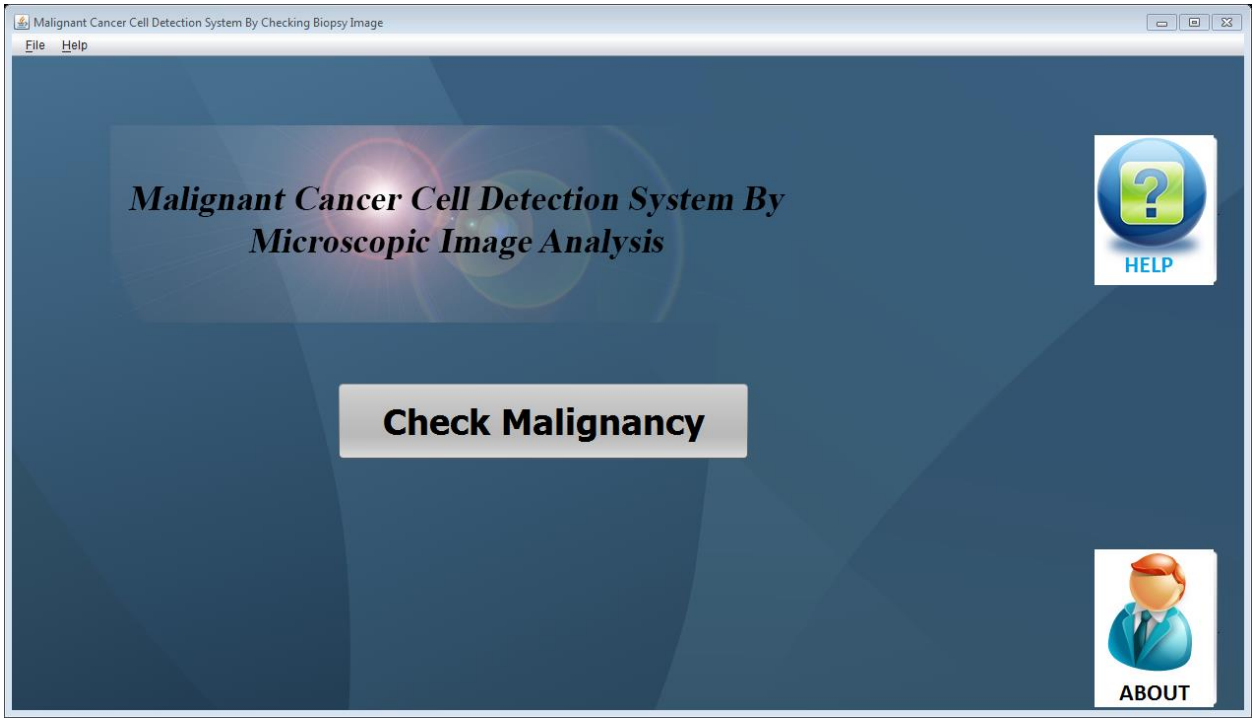

Fig 4.1: User Interface GUI

In the interface we give buttons for Help and about button for helping the user and the main button is **Check Malignancy** button.

TH e **Help Button** will help to use the software. **About Button** will help to know about the developers. And the main functional button is **Check Malignancy Button**. If anyone

want to check the malignancy of a cell need to click this button then a new window will be opened. On that window there is a button **Browse.** Click the browse button to browse the file that you want to check malignancy.

The following figure shows the browse window:

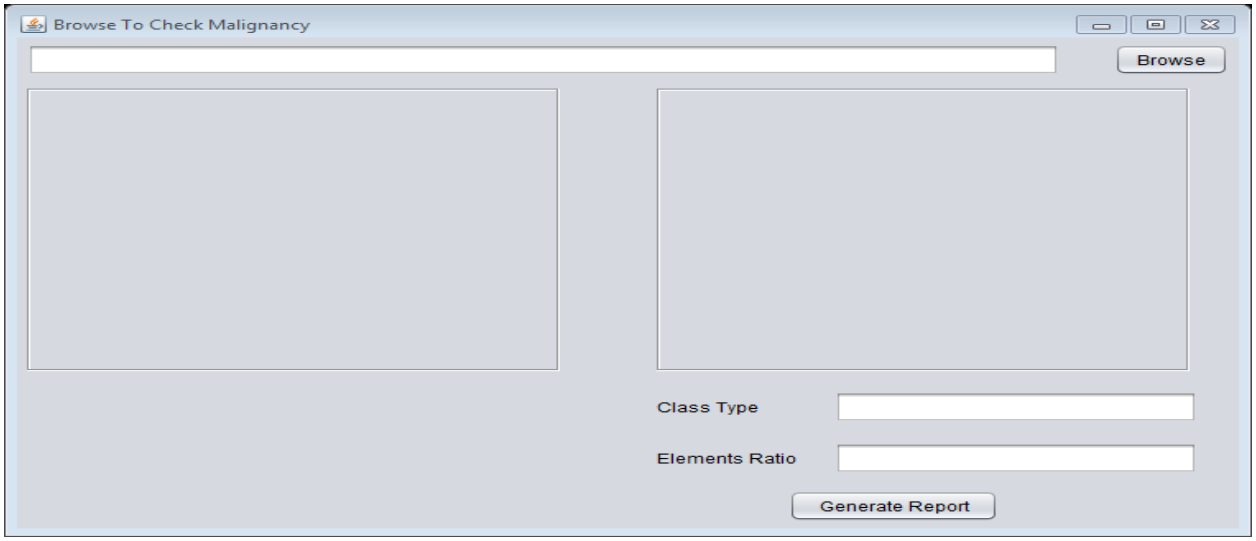

Fig 4.2: Browse Window GUI

When file is browsed it will be checked and result will be shown in which it shows the Class type and Elements ratio. The window will look like the following;

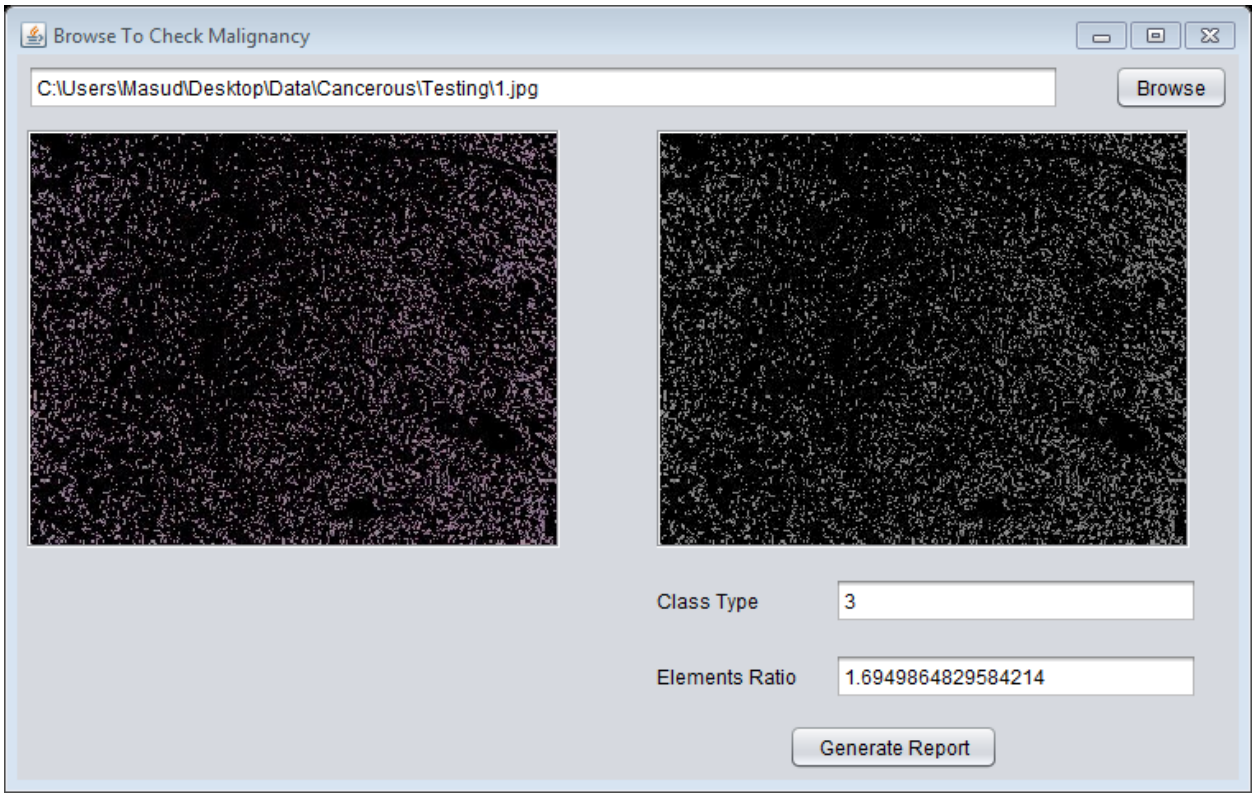

Fig 4.3: Showing testing Result

After that if want to generate report click the **Generate Report Button.** Then report will be generated. Its look like the figure in the next page named Report View.

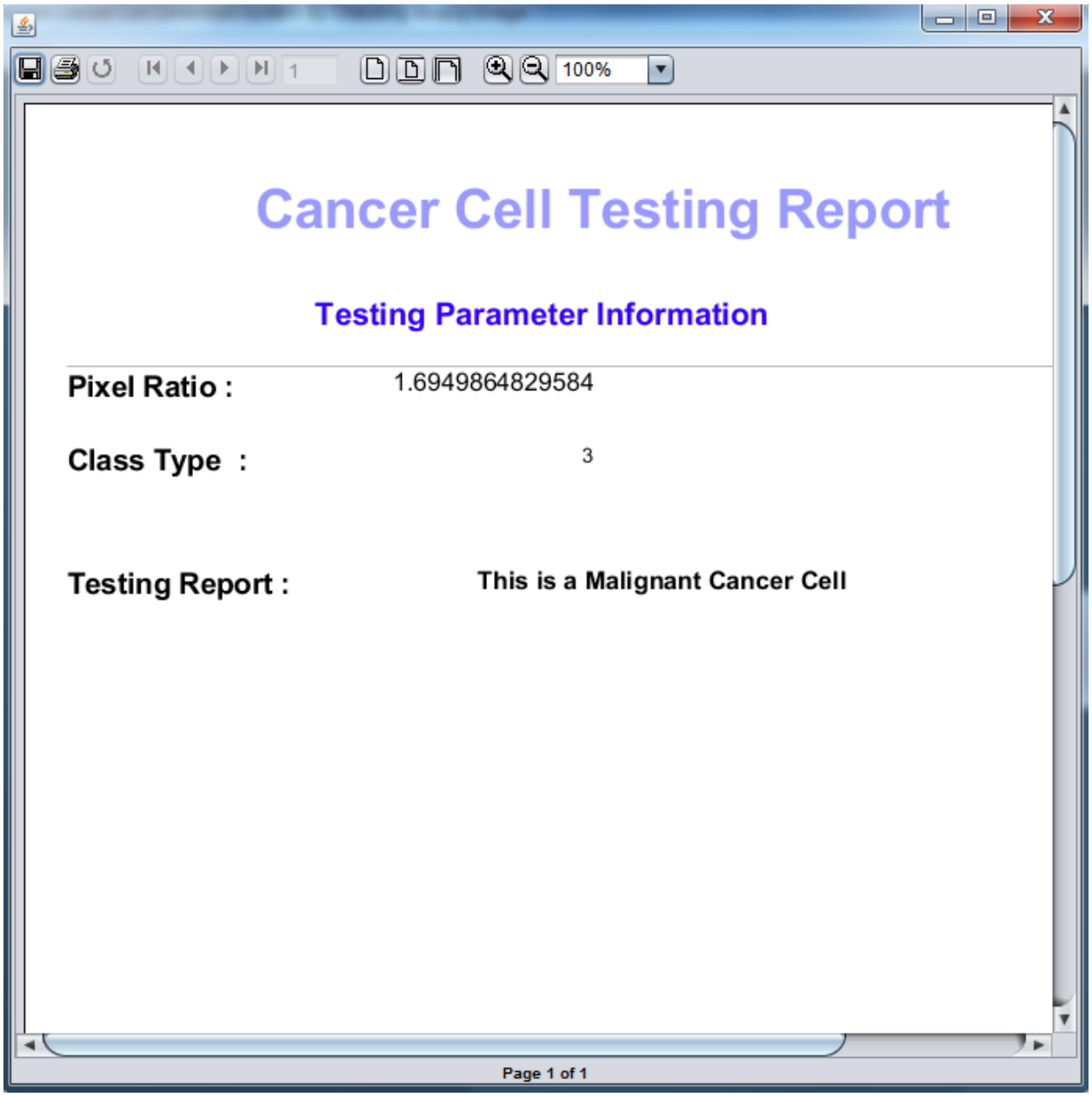

Fig 4.4: Report View.

If needed report can be printed.

4.2 Testing

The most challenging task of this work was Data collection. It was not possible to collect data from local diagnostic center as cause of their security or other administrative reason. At last data are found that was given with a paper in the internet by contacting with the publisher of the paper. The name of the paper is "Breast Cancer Diagnosis From Biopsy Images Using Generic Features and SVMs" by Brook et all.

In the data set there are three categories of data first is Normal image cell, second is insitu cell image and invasive cell image. The main characteristics of the images is that they are 60% zoomed.

MATLAB software package version 10 is used to implement the software in the current work. When the training process is completed for the training data (150 cases).The testing process is done for 50 cases. These 100 cases are fed to the proposed network and their output is recorded. Table I list the data division of Breast cancer disease data. Summary of data division is described in Table I. Three classes of tumor divisions came from three attributes with continuous values. The samples of data were categorized into three classes.

Table 4.1: Data Division of Cancer Detection

Class	Training	Validation	Testing
Class 1: Normal Cell	25	15	10
Class 2: Tumorous Cell	23	15	13
Class 3: Cancer Cell	23	15	13
Total Data	96	45	33

The proposed image processing method is implemented as an algorithm tested by MATLAB simulation with the aid of experimental breast H & E stained histopathology image data. The accuracy of classifier is defined as the ratio of the number of samples correctly classified to the total number of samples tested. The trained system has been

tested in the retrieval mode, in which the testing vectors are not taking part in the training process. We are used the standard multilayered feed forward back propagation neural network algorithms.

sTable 4.2: Performance Result

Case Study	Training	Validation	Testing
Class 1	95.00	93.10	93.50
Class 2	95.00	93.30	93.80
Class 3	95.50	93.50	94.00

CHAPTER 5: CONCLUSION AND FUTURE SCOPE

Cancer has become one of the biggest threats to human life, and is expected to become the leading cause of death over the next few decades. Based on the statistics of World Health Organization (WHO), cancer accounted for 13% of all deaths in 2005 worldwide. Deaths due to cancer are expected to rise in the future, with an estimated 11.4 million deaths by 2030 with 10-25% of the cancer cells going unrecognized by radiologist due to non-accurate diagnosis. The manual diagnosis process is so much time consuming. As for that a faster way of detection is necessary.

Nowadays a number of works are going on in for faster cancer detection, mostly works are happening on image processing arena. Our work, Cancer Detection is also in image processing field. We have used image processing techniques for analyzing the image and understanding its features.

In the beginning of the work we studied many works already done in this field and tried to find out weakness of some of them.

The working steps in our proposed technique are transferring image color space, color based segmentation and clustering, segmenting blue nuclei, feature extraction and decision making. Through these step we do our work. In first step we convert the main image into L*a*b color space then we do color based segmentation of the image. Clustering is done after the segmentation process. As we use k-means and define n is equal to 3 so the number cluster is 3. Among the cluster we manually selected the cluster that is seems to be more reliable to extract our necessary features like nucleus, black pixel and pixel ratio etc. Then we ran some procedure on that and highlight the blue nuclei. After that, the image is used in the further process that occurred in our developed software. We developed the software in java technology. We trained the develop system by a number of data so that it will as take input the found image and the internal works are like, first the system will convert the image in binary, then count the number of connected components which are the blue nuclei, after that it counts black and white pixel numbers and their ratio. These three will be provided as parameters and system will compare the value and return the class which it would be belongs to.

5.1 Future Scope

As our work is on cancer detection, which is a vast field of research. Lots of improvement and further developments can be performed on our method.

If we talk about our system, it is a semi-automated system. We need to make it semi-automated because we had to select cluster manually. Our first future work is to make the system properly that it would be able to select the cluster its own. There is also some weakness in our segmentation and noise removal algorithm, fixing it also is a part of our future work. The decision making this system, we have only trained it and tested it. Another future work is making it capable of learning in each testing case and also including self-accuracy check system.

This work not only providing some future work, it also creates many working scope for us. As our system detects the cancerous cell by testing a biopsy image, this system includes many procedures like color based segmentation, clustering and feature extraction which provides a vast field for working. With the help of this technique, several works can be done on **pattern recognition**, **Face detection or character identification**, **Visual Inspection** and **Recognition** also can be done. With the process of some normal work like feature extraction many information from a single image can be extracted, by using connected component counting any one can count the number of specified components in an image. In a word through this system each of the technique in this work has many working field. In the field of **medical science** it has many scope of work and the topics of cancer detection is also a topic of rapidly growing interest and applicability.

REFERENCES

[1]Demir et al. , "Automated cancer diagnosis based on histopathological images: a systematic survey",TECHNICAL REPORT, RENSSELAER POLYTECHNIC INSTITUTE, DEPARTMENT OF COMPUTER SCIENCE, TR-05-09.

[2]Shekhar et al. ,“Breast Cancer Detection and Classification of Histopathological Images”. International Journal of Engineering Science and Technology (IJEST)

[3]Lei et al. , “Computer Assisted Diagnosis in Histopathology”, National Library of Medicine National Institutes of Health, USA

[4]Swarnalatha et al. ,“Microscopic Image Segmentation and Recognition on Cancerous Cells”, International Journal of Engineering Science and Technology, Vol. 2(11), 2010, 6771-6778

[5] Yin et al. ,“Microscopic Image Analysis and Recognition On Pathological Cells", IEEE, 2007

[6]seer.cancer.govstatfactshtmlall.html

[7]www. who. int/ hpvcentre

[8]L Zhang and Q G Wang, "Processing Technology in Microscopic Images of Cancer Cells in Pleural Fluid Based on Fuzzy Edge Detection Method", Journal of Physics: Conference Series 48 (2006) 329–333

[9]Shekhar Singh,“Cancer Cells Detection and Classification in Biopsy Image”, International Journal of Engineering Science and Technology (IJEST)

[10]Alexander et al. ,"Breast Cancer Diagnosis From Biopsy Images Using Generic Features and SVMs".

[11]GEOFF DOUGHERTY, Digital Image Processing for Medical Applications,Publisher Cambridge University Press, page 47, Cambridge, 2009

[12]GEOFF DOUGHERTY, Digital Image Processing for Medical Applications,Publisher Cambridge University Press, page 90, Cambridge, 2009

[13]Rich And Knight, Artificial Intelligence, Tata McGraw-Hill, Chapter 17,2nd edition

[14]Rich And Knight, Artificial Intelligence, Tata McGraw-Hill, Chapter 18,2nd edition

[15]MICHAEL NEGNEVITSKY, Artificial Intelligence, Pearson Publication, Chapter 6,2nd edition 2005

[16]Simon & Hakin, Neural Networks A comprehasive Foundation,Pearson Publication, Chapter 1,Eight Edition 2005

[17]Simon & Hakin, Neural Networks A comprehasive Foundation,Pearson Publication, Chapter 2,Eight Edition 2005

[18] Micheline Kamber, Data Mining Concepts and Techniques, Diane Cerra, Chapter 5, Second Edition

[19] CH Yip, "Breast Cancer In Asia", Technical Report, Faculty Of Medicin University Of Malay