Sudan University of Science and Technology

College of Graduate Studies and Scientific Research

Detection of breast cancer using artificial neural networks (ANN)

تشخيص سرطان الثدي باستخدام الشبكات العصبية الاصطناعية

Dissertation submitted in partial fulfillment for the degree of Master of Science in biomedical engineering

By: Toga Farag Eltayeb

Supervised by: Dr.Eltahir Mohamed Hussein

June 2014

Dedication

I dedicate this work firstly to my father and to my best sister”Nawal”, gone now but never forgotten, and I will miss them always and love them forever.

Lovingly I dedicate this thesis to my wonderful mother. Her measureless support, encouragement and constant love have sustained me throughout my life. Without her caring support it would not have. I hope from Allah to keep her and give her the health and force to continue her message in this life.

I dedicated also to my husband who helped and support me in each step of the way. And I dedicated for my sisters and brothers. Finally yet importantly I dedicate this project for every one that helped me to be at the place that I am today.

Acknowledgement

I would like to express my deep thanks to Allah and then I have to thank my mother for her love and support throughout my life. Thank you for everything. And I would like to thanks my husband for endless helps in all my way.

Further I would like to sincerely thank my supervisor, ”Dr.Eltahir Mohamed Hussein”, for his patience, guidance and support throughout this study, and especially for his confidence in me.

Finally I would like to thank my teacher and my friend “Fatima”, who always had something nice to say and who supporting me through the more difficult times of my life.

Abstract

The aim of this project is to design and implement a Matlab based image processing software to extract features of breast cancer images in order to classify disease through neural network in order to examine system accuracy by comparing two techniques of neural networks, feed-forward and back-propagation neural networks. After simulation it was found that the feed forward neural network results are approximately same as the results of back propagation neural network, but from the accuracy side the feed forward neural network (94.11%) is better than the back propagation neural network (92%).

المستخلص

الهدف من هذا البحث إنشاء منظومة في Matlab لمعالجة واستخلاص مواصفات صورة سرطان الثدي وتصنيف هذا المرض من خلال الشبكات العصبية الاصطناعية لاختبار دقة هذا النظام بمقارنة نوعين من الشبكات العصبية feed forward والـ back propagation.

بعد التمثيل , وجد أن نتائج شبكة feed forward مشابهة تقريباً لنتائج شبكة الـ back propagation. ولكن من ناحية الدقة تقدمت شبكة ال feed forward (94.11%) على شبكة ال(92%)back propagation.

Contents

List of figures

List of tables

Abbreviation

ANN	Artificial Neural Network
GLCM	Gray scale Level Co-occurrence Matrix
BP	Back Propagation
GA	Genetic Algorithm
CGP	Cartesian Genetic Programming
FNA	Fine Needle Aspiration
LVQ	Learning Vector Quantization
SE	Sensitivity
SP	Specifity
AC	Accuracy
MCC	Mathews Correlation Coefficient
FFANN	Feed Forward Artificial Neural Network
BPANN	Back Propagation Artificial Neural Network

Chapter One

Introduction

1.1 General View

Breast cancer is the second most lethal cancer for women in the world today. Early detection of the breast cancer can reduce mortality rate. The breast Tumor are of two types first is Benign and second is Malignant, Benign Tumor which is non-cancerous and not life threatening and Malignant Tumor are cancerous and life threatening. Detection of breast cancer can be achieved using Digital Mammography and classification of breast tumor whether is Malignant or Benign. This thesis presents a research on breast cases of 75 samples proven breast tumor are analyzed and classified into benign and malignant categories using ANN. Two different feature extraction methods used here are Gray scale Level Co-occurrence Matrix (GLCM) and Intensity based features. A supervised classifier system based on neural network is used. The performance of the each feature extraction method is evaluated on Digital mammogram images. The experimental results suggest that GLCM method outperformed the other one.

Breast cancer is the most frequent cancer in women worldwide. The disease is curable if detected early enough. Screening is carried out on the basis of mammograms, which use x-ray images to reveal lumps in the breast. Calcium deposits can also indicate the existence of a tumor. However, the deposits are often only a few tenths of a millimeter in size and so deeply embedded in dense tissue that they are nearly undetectable in the images. Mammogram samples with marked malignant tumor as shown in figure (1.1). Digital mammography are proven as efficient tool to detect breast cancer before clinical symptoms appear [11].

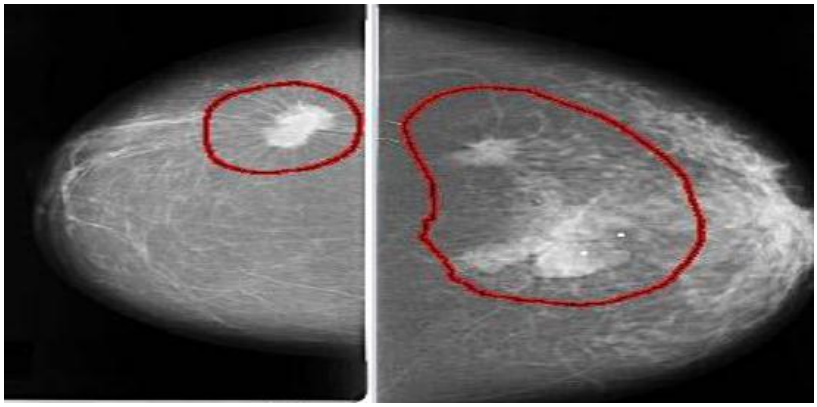


Figure (1.1) Mammogram samples with marked malignant tumor

Digital mammography is currently considered as standard procedure for breast cancer diagnosis, various artificial intelligence techniques are used for classification problems in the area of medical diagnosis. Feature extraction of image is important step in mammogram classification. These features are extracted using image processing techniques. Several types of feature extraction from digital mammograms including position feature, shape feature and texture feature etc. Textures are one of the important features used for many applications. Texture features have been widely used in mammogram classification. The texture features are ability to distinguish between abnormal and normal cases. Texture can be characterized as the space distribution of gray levels in a neighbourhood [1, 2]. Texture feature have been proven to be useful in differentiating normal and abnormal pattern. Extracted texture features provide information about textural characteristics of the image. Different classifiers are used for medical imaging application including artificial intelligence, wavelet etc. Texture measures are two types, first order and second order. In the first order, texture measure are statistics calculated from an individual pixel and do not consider pixel neighbor relationships. Intensity feature are first order texture calculation. In the second order, measures consider the relationship between neighbor pixels GLCM is a

second order texture calculation [3, 4]. Texture features has been extracted and used as parameter to enhance the classification result.

1.2 Problem Statement

The mammogram is the device used for breast cancer detection, the device produce image with high resolution and with a size of 1024 by 1024, these images affected by noise that reduce the quality of detection.Using manual detection of breast cancer can result Low accuracy detection and diagnosis.

Each imaging device has a calibration which can vary between devices so each image has a unique contrast depend on device type or device configurations.

1.3 Objectives

1. Using Matlab as an image processing tool to extract features of breast cancer images in order to classify disease.
2. To use an artificial neural network and examine system accuracy by comparing two techniques of neural networks, feed-forward and back-propagation.

1.4 Methodology

Using matlab software and the image processing techniques available by the software for breast cancer detection moreover the RawData used in this project is about 75 sample including normal and abnormal, all of the raw data entered to a preprocessing stage in order to be viewed and to be filtered from noise, then adjusting contrast of image, the segmentation progress is applied in order to detect tumor cells , after that feature extracted for classification ,then two types of neural networks was used for classification and to compare between their accuracy.

1.5 literature review

Afzan Adam &Khairuddin Omar [1]

Breast cancer has become a common mortality factor in Malaysia. Despite the fact, not all general hospitals have the mammogram facilities. The long waiting for diagnosing a breast cancer may increase the possibility of fatality and the cancer spreading. Therefore a computerized breast cancer diagnosis prototype has been developed to reduce the time taken and indirectly reducing the probability of death. As Back Propagation Neural Network (BPNN) is commonly used in medical field [9], the prototype model will use this method as a classifier as well. However, BPNN has several issue and weakness to be address. Therefore, the architecture will be enhanced with Genetic Algorithm (GA) technology. By combining the GA and BPNN, a faster classifier model can be developed, without downgrading the classification performance. The development of this model, namely GAwNN; is aimed to reduce the diagnosis time as well as increasing the accuracy percentage in classifying mass in breast to either benign, or malignant.

Dr. N.GanesanDr.K.VenkateshDr.N.AramaA.MalathiPalani[2]

Artificial Neural Network is a branch of Artificial intelligence, has been accepted as a new technology in computer science. Neural Networks are currently a 'hot' research area in medicine, particularly in the fields of radiology, urology, cardiology, oncology and etc. It has a huge application in many areas such as education, business; medical, engineering and manufacturing .Neural Network plays an important role in a decision support system. In this paper, an attempt has been made to make use of neural networks in the medical field (carcinogenesis (preclinical study)). In carcinogenesis, artificial neural networks have been successfully applied to the problems in both pre-clinical and post-clinical diagnosis. The main aim of research in medical diagnostics is to

develop more cost-effective and easy–to-use systems, procedures and methods for supporting clinicians. It has been used to analyze demographic data from lung cancer patients with a view to developing diagnostic algorithms that might improve triage practices in the emergency department. For the lung cancer diagnosis problem, the concise rules extracted from the network achieve a high accuracy rate of on the training data set and on the test data set.

ArbabMasood Ahmad, GulMuhammd Khan, S.Ali Mahmud [3]

A fast learning neuro-evolutionary technique that evolves Artificial Neural Networks using Cartesian Genetic Programming (CGPANN) is used to detect the presence of breast cancer. Features from breast mass are extracted using fine needle aspiration (FNA) and are applied to the CGPANN for diagnosis of breast cancer. FNA data is obtained from the Wisconsin Diagnostic Breast Cancer website and is used for training and testing the network. The developed system produces fast and accurate results when compared to contemporary work done in the field. The error of the model comes out to be as low as 1% for Type-I (classifying benign sample falsely as malignant) and 0.5% for Type- II (classifying malignant sample falsely as benign).

1.6Thesis layout

This thesis consist of six chapters,Chapter One is an introduction, Artificial Neural Network technique andpresented in chapter two, the anatomy of breast is discussed in chapter three, chapter four describe the proposed system(methodology), while chapter five represent results and discussion finally the conclusion and recommendation represented in chapter six.

Chapter Two

Artificial Neural Network technique

2.1 Artificial neural network

An artificial neural network (ANN) is an information processing system which contains a large number of highly interconnected processing neurons. These neurons work together in a distributed manner to learn from the input information, to coordinate internal processing, and to optimize its final output. In figure (2.1), each circular node represents an artificial neuron and an arrow represents a connection from the output of one neuron to the input of another.

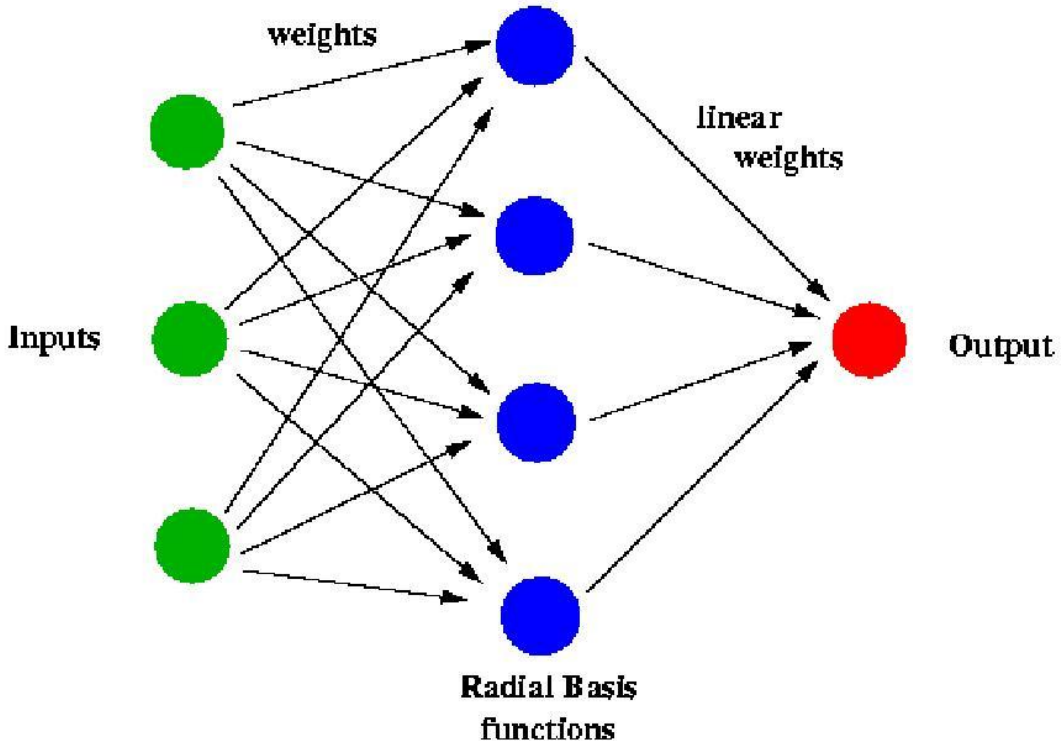


Figure (2.1) example of Neural Network

In computer science and related fields, artificial neural networks are computational models inspired by animals' central nervous systems (in

particular the brain) that are capable of machine learning and pattern recognition. They are usually presented as systems of interconnected "neurons" that can compute values from inputs by feeding information through the network.

For example, in a neural network for handwriting recognition, a set of input neurons may be activated by the pixels of an input image representing a letter or digit. The activations of these neurons are then passed on, weighted and transformed by some function determined by the network's designer, to other neurons, etc., until finally an output neuron is activated that determines which character was read.

Like other machine learning methods, neural networks have been used to solve a wide variety of tasks that are hard to solve using ordinary rule-based programming, including computer vision and speech recognition.

The inspiration for neural networks came from examination of central nervous systems. In an artificial neural network, simple artificial nodes, called "neurons", "neurodes", "processing elements" or "units", are connected together to form a network which mimics a biological neural network.

There is no single formal definition of what an artificial neural network is. Commonly, though, a class of statistical models will be called "neural" if they

consist of sets of adaptive weights, i.e. numerical parameters that are tuned by a learning algorithm, and are capable of approximating non-linear functions of their inputs.

The adaptive weights are conceptually connection strengths between neurons, which are activated during training and prediction.

Neural networks are also similar to biological neural networks in performing functions collectively and in parallel by the units, rather than there being a clear delineation of subtasks to which various units are assigned. The term "neural network" usually refers to models employed in statistics, cognitive psychology and artificial intelligence. Neural network models which emulate the central nervous system are part of theoretical neuroscience and computational neuroscience.

In modern software implementations of artificial neural networks, the approach inspired by biology has been largely abandoned for a more practical approach based on statistics and signal processing. In some of these systems, neural networks or parts of neural networks (like artificial neurons) form components in larger systems that combine both adaptive and non-adaptive elements. While the more general approach of such systems is more suitable for real-world problem solving, it has far less to do with the traditional artificial intelligence connectionist models. What they do have in common, however, is the principle of non-linear, distributed, parallel and local processing and adaptation. Historically, the use of neural networks models marked a paradigm shift in the late eighties from high-level (symbolic) artificial intelligence, characterized by expert systems with knowledge embodied in if-then rules, to low-level (sub-symbolic) machine learning, characterized by knowledge embodied in the parameters of a dynamical system.

2.2 Types of artificial neural networks

There are many types of artificial neural networks (ANN). An artificial neural network is a computational simulation of a biological neural network. These models mimic the real life behavior of neurons and the electrical messages they produce between input (such as from the eyes or

nerve endings in the hand), processing by the brain and the final output from the brain (such as reacting to light or from sensing touch or heat). There are other ANNs which are adaptive systems used to model things such as environments and population.

The systems can be hardware and software based specifically built systems or purely software based and run in computer models.

2.2.1 Feedforward neural network

The feed forward neural network was the first and arguably most simple type of artificial neural network devised. In this network the information moves in only one direction — forwards: From the input nodes data goes through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. Feed forward networks can be constructed from different types of units, e.g. binary McCulloch-Pitts neurons, the simplest example being the perceptron. Continuous neurons, frequently with sigmoid activation, are used in the context of back propagation of error.

2.2.2 Radial basis function (RBF) network

Radial basis functions are powerful techniques for interpolation in multidimensional space. A RBF is a function which has built into a distance criterion with respect to a center. Radial basis functions have been applied in the area of neural networks where they may be used as a replacement for the sigmoid hidden layer transfer characteristic in multi-layer perceptrons. RBF networks have two layers of processing: In the first, input is mapped onto each RBF in the 'hidden' layer. The RBF chosen is usually a Gaussian. In regression problems the output layer is then a linear combination of hidden layer values representing mean predicted output. The interpretation of this output layer value is the same

as a regression model in statistics. In classification problems the output layer is typically a sigmoid function of a linear combination of hidden layer values, representing a posterior probability. Performance in both cases is often improved by shrinkage techniques, known as ridge regression in classical statistics and known to correspond to a prior belief in small parameter values (and therefore smooth output functions) in a Bayesian framework.

2.2.3 Kohonen self-organizing network

The self-organizing map (SOM) invented by TeuvoKohonen performs a form of unsupervised learning. A set of artificial neurons learn to map points in an input space to coordinates in an output space. The input space can have different dimensions and topology from the output space, and the SOM will attempt to preserve these.

Learning Vector Quantization (LVQ) can also be interpreted as a neural network architecture. It was suggested by TeuvoKohonen, originally. In LVQ, prototypical representatives of the classes parameterize, together with an appropriate distance measure, a distance-based classification scheme.

2.2.4 Recurrent neural network

Contrary to feed forward networks, recurrent neural networks (RNNs) are models with bi-directional data flow. While a feed forward network propagates data linearly from input to output, RNNs also propagate data from later processing stages to earlier stages. RNNs can be used as general sequence processors. [5]

Chapter Three

Anatomy of breast

3.1 Anatomy and physiology of the breast

Women and men both have breasts, but women have more breast tissue than men. Each breast lies over a muscle of the chest called the pectoral muscle. The female breast covers a fairly large area. It extends from just below the collarbone (clavicle), to the armpit (axilla) and across to the breastbone (sternum) as shown in figure (3.1).

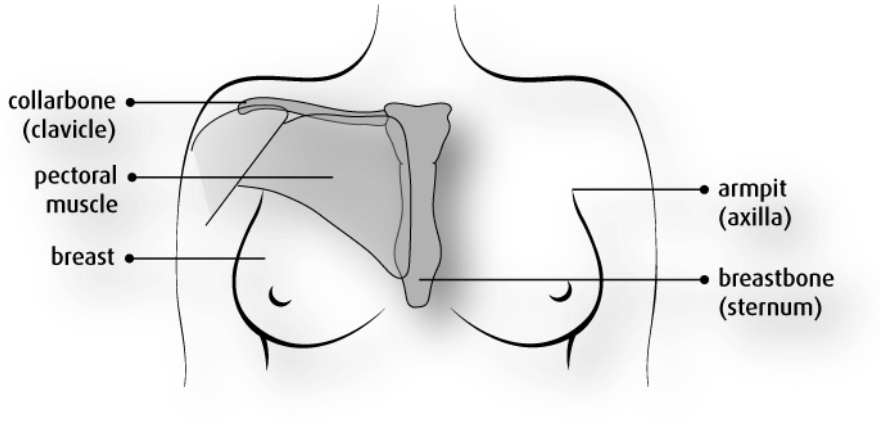


Figure (3.1) the breast and surrounding structure

3.1.1 Structure of breast

The breast is a mass of glandular, fatty and connective tissue. As shown in figure (3.2) the breast is made up of:

- lobules – glands that produce milk
- ducts – tubes that carry milk from the lobules to the nipple
- fatty and connective tissue – surrounds and protects the ducts and lobules and gives shape to the breast

- areola – the pink or brown, circular area around the nipple that contains small sweat glands, which release (secrete) moisture as a lubricant during breast-feeding
- nipple – the area at the center of the areola where the milk comes out

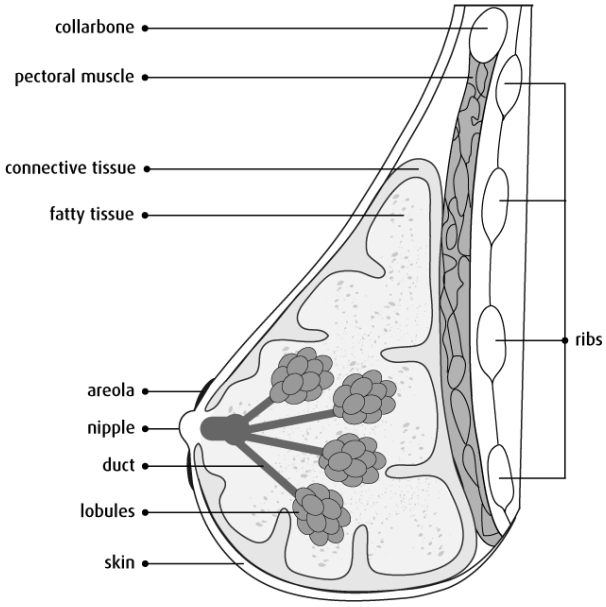


Figure (3.2) the Breast

Ligaments support the breast. They run from the skin through the breast and attach to muscles on the chest.

There are several major nerves in the breast area, including nerves in the chest and arm. There are also sensory nerves in the skin of the chest and axilla.

3.1.2 The lymphatic system of the breast

The breast has many blood vessels and lymph vessels. Lymph vessels are thin tubes similar to blood vessels. They collect and move lymph fluid away from the breast into small bean-shaped masses of lymphatic tissue, called lymph nodes, in the area around the breast. The lymph vessels and

lymph nodes are part of the lymphatic system, which helps fight infections.

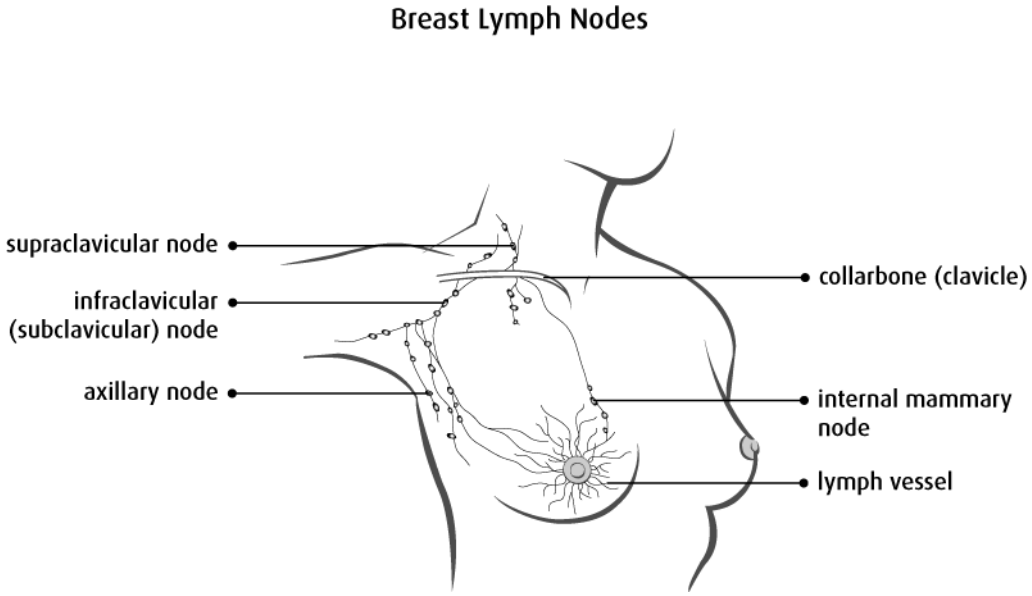


Figure (3.3) Breast Lymph Nodes

As shown in figure (3.3) the breast lymph nodes include:

- supraclavicular nodes – above the collarbone
- infraclavicular (or subclavicular) nodes – below the collarbone
- axillary nodes – in the armpit (axilla)
- internal mammary nodes – inside the chest around the breastbone (sternum)

3.1.3 Axillary lymph nodes

There are about 30–50 lymph nodes in the axilla. The number varies from woman to woman.

The axillary lymph nodes are divided into 3 levels as shown in figure (3.4) according to how close they are to the pectoral muscle on the chest:

- **level I (low axilla) –** located in the lower or bottom part of the armpit, along the outside border of the pectoral muscle

- **level II (mid axilla) –** located in the middle part of the armpit, beneath the pectoral muscle
- **level III (high axilla) –** located below and near the center of the collarbone, above the breast area and along the inside border of the pectoral muscle

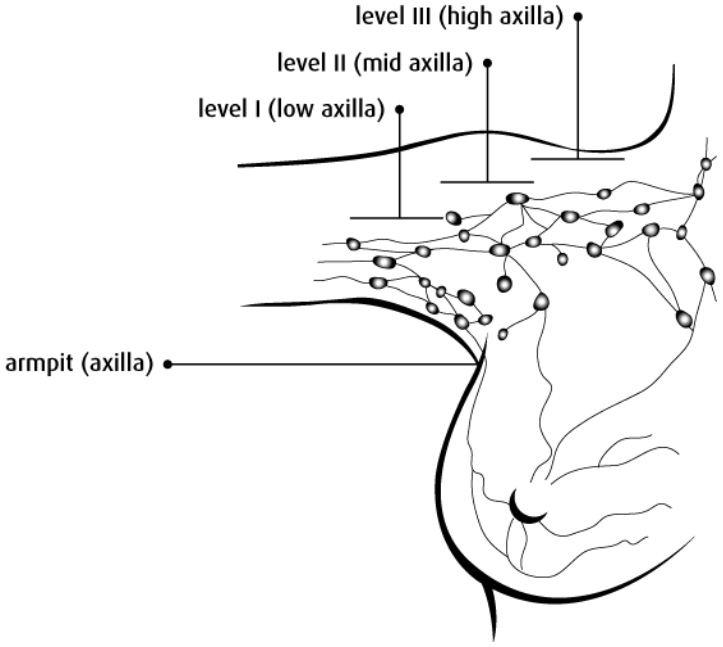


Figure (3.4) Axillary Lymph Nodes

When breast cancer spreads, it usually goes to level I lymph nodes first, to level II next and then to level III.

3.1.4 Breast development

Breast tissue changes at different times during a woman's life. It changes during puberty, during the menstrual cycle, during pregnancy and after menopause.

Female breasts do not begin growing until puberty (around 10–12 years of age). At this time, the breasts respond to hormonal changes (mostly increased estrogen and progesterone) in the body and begin to develop. During puberty, the breast ducts and milk glands grow. The breast skin

stretches as the breasts grow, creating a rounded appearance. Young women tend to have denser breasts (more glandular tissue) than older women.

In older women, much of the glandular and ductal tissue is replaced with fatty tissue and breasts become less dense. Ligaments also lose their elasticity when women age, causing the breasts to sag.

The size and shape of women's breasts vary considerably. Some women have a large amount of breast tissue and have larger breasts. Others have a smaller amount of tissue with little breast fat. A woman's breasts are rarely the same size. Often one breast is slightly larger or smaller, higher or lower or shaped differently than the other.

3.1.5 Hormones and the breast

Estrogen is the main female hormone. It influences female sexual characteristics, such as breast development, and it is necessary for reproduction. Most of the estrogen in a woman's body is made by the ovaries, though a small amount is made by the adrenal glands.

Progesterone is the other female sex hormone made in the ovaries. Its role is to prepare the uterus (womb) for pregnancy and the breasts for producing milk for breast-feeding (lactation).

The breast tissues are exposed to monthly cycles of estrogen and progesterone throughout a woman's childbearing years.

- In the first part of the menstrual cycle, estrogen stimulates the growth of the milk ducts.
- Progesterone takes over in the second part of a woman's menstrual cycle, stimulating the lobules.

After menopause, the monthly cycle of estrogen and progesterone end. However, the adrenal glands continue to produce estrogen so that a woman keeps her sexual characteristics.

3.1.6 Function of breast

The breast's main function is to produce, store and release milk to feed a baby. Milk is produced in lobules throughout the breast when they are stimulated by hormones in a woman's body after giving birth. The ducts carry the milk to the nipple. Milk passes from the nipple to the baby during breast-feeding. [12]

Chapter Four
The proposed system

4.1 System Block Diagram

The system is consist of preprocessing block that filter image from noise and adjust the contrast to insure that the image is ready to be entered to the feature extraction , from the extracted features a comparison between these features is applied to classify the breast cancer type the diagnosis results then is displayed.

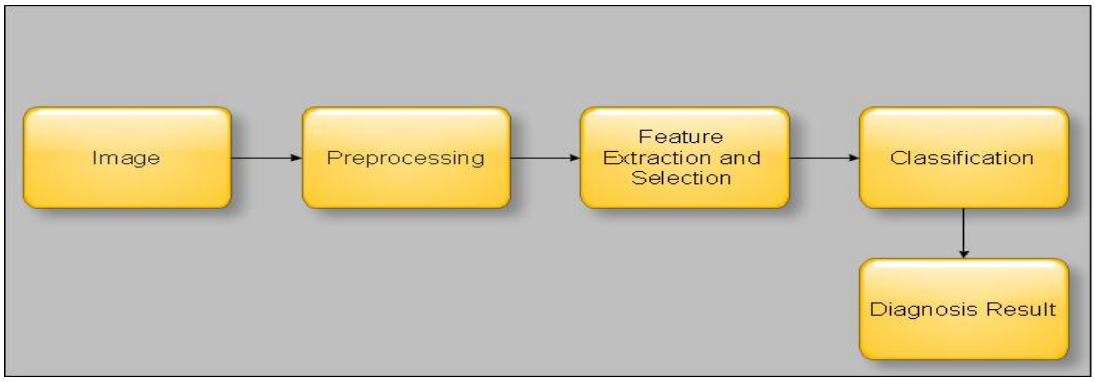


Figure (4.1) system block diagram

4.2 Block diagram description

4.2.1 Image:

High resolution image from mammography device, images of Mammogram are available in the Department of Electrical Engineering, MIAS Gwalior. These images are available with the same specification (3000x4500 pixels with 16-bit pixel depth) and the total numbers of images are 75 in which 25 are healthy cases and remaining are diseased cases.

4.2.2 Preprocessing:

In this stage the image will be loaded into Matlab and a noise will be removed, then the median filter then adaptive median filter, contrast limited adaptive histogram equalization filter, finally the segmentation was done.

Preprocessing steps

a)load original image:the original image shown in figure (4.2) will be loaded into matlab**.**

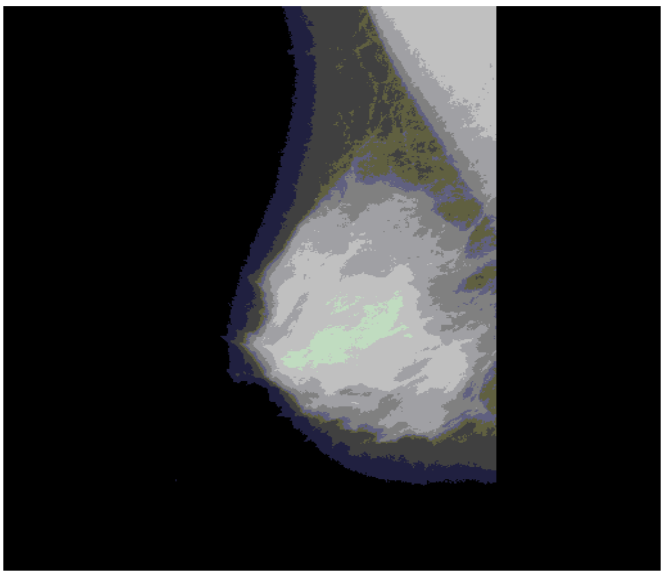


Figure (4.2) original image

b) Apply median filter

In the median filter the image is converted into grayscale in order to adjust the overall color variation and remove the interferance between colors with a fixed scale asshown in figure(4.3).

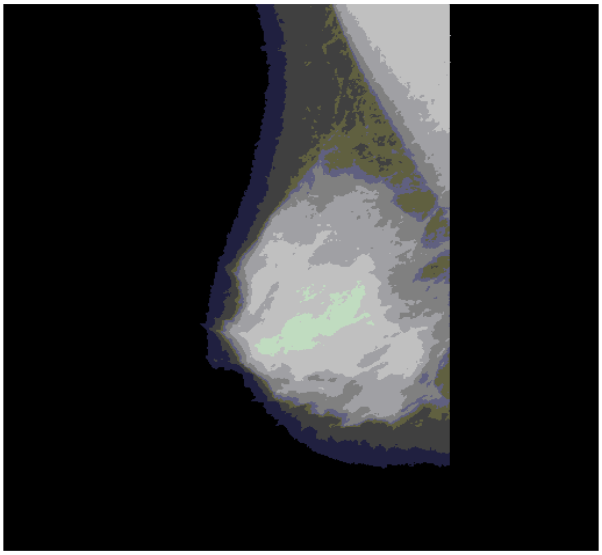


Figure (4.3) Apply median filter

c) Apply adaptive median filter

This filter use to detect each part of the image by creating a scanning matrix 3x3 in order to find the high medium and low pixel color in order to adjust the matirix to the medimum of the colors as shown in figure(4.4).

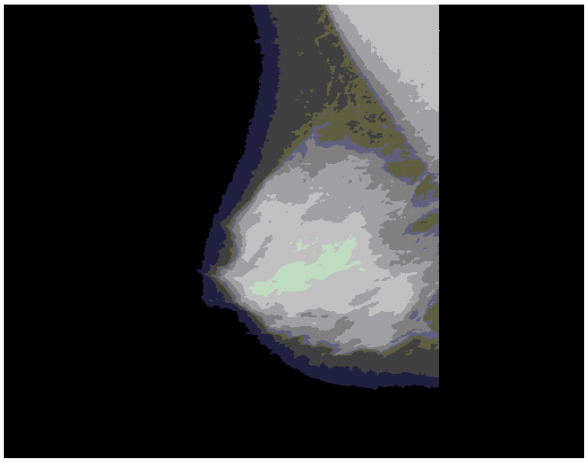


Figure (4.4) Apply adaptive median filter

d) Contrast limited adaptive histogram equalization

The Contrast limited adaptive histogram equlization used To adjust the edge of image to have the contrast value of the pixel contrast as shown in figure(4.5).

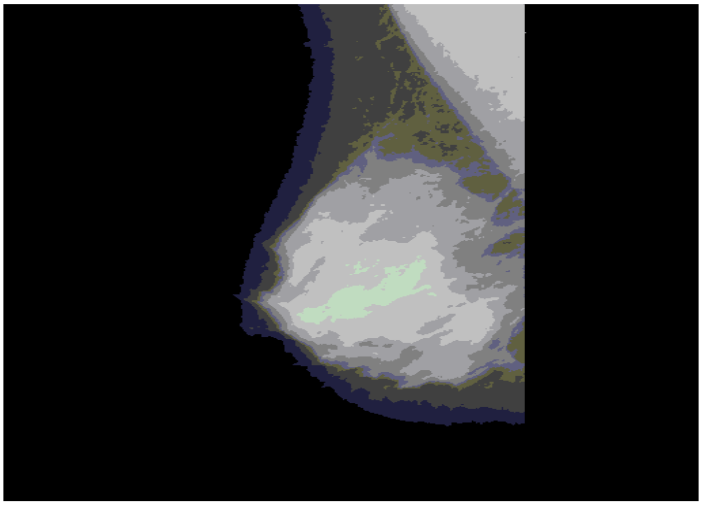


Figure (4.5) Contrast limited adaptive histogram equalization

f) Segmentation

the separate each color from the image in a seperated layer as shown in figure(4.6).

Figure (4.6) result image from segmentation

Infected area Detection Steps

Step 1: Read Image

Read in the cell.tif image as shown in figure(4.7), which is an image of a breast cancer cell.

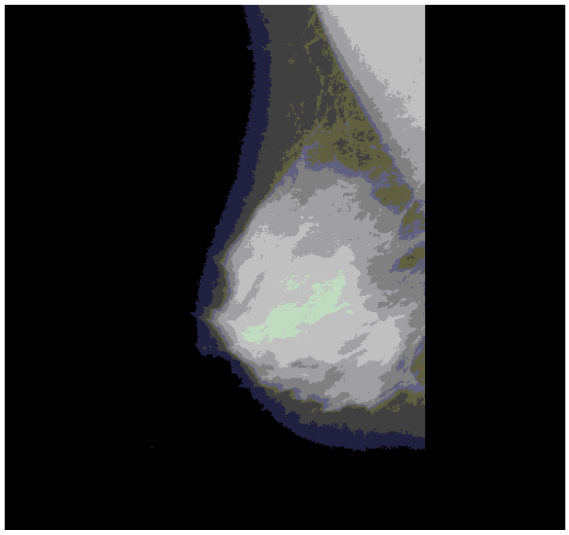


Figure (4.7) loading original image

Step 2: Detect Entire Cell

Only infected area will be detected using a technieques of object detection by applying segmentation. The object to be segmented differs greatly in contrast from the background image. Changes in contrast can be detected by operators that calculate the gradient of an image. The gradient image can be calculated and a threshold can be applied to create a binary mask containing the segmented cell. First, the usage edge and the Sobel operator to calculate the threshold value. then a tuning process done uisng the threshold value and use edge again to obtain a binary mask that contains the segmented cell as shown in figure(4.8).

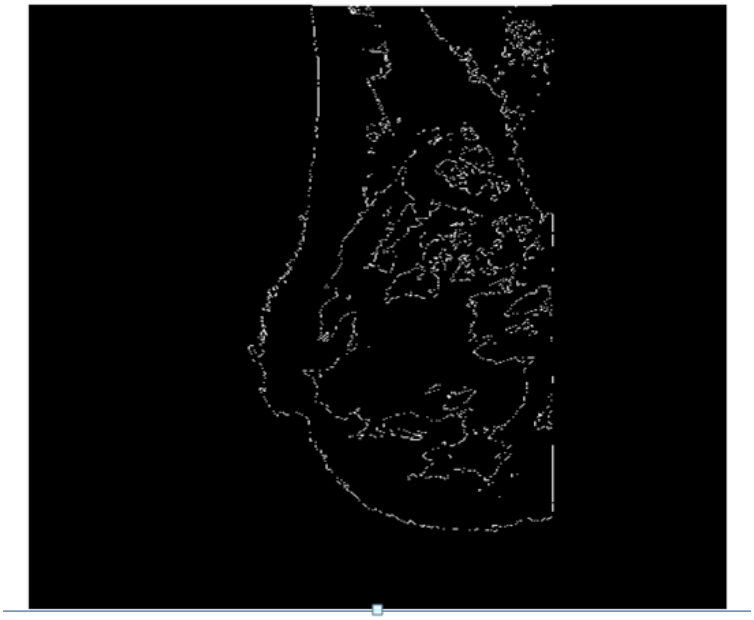


Figure (4.8) Image binary gradient mask

Step 3: Dilate the Image

The binary gradient mask shows lines of high contrast in the image as shown in figure(4.9). These lines do not quite delineate the outline of the object of interest. Compared to the original image, you can see gaps in the lines surrounding the object in the gradient mask. These linear gaps will disappear if the Sobel image is dilated using linear structuring elements, which we can create with the strel function.

The binary gradient mask is dilated using the vertical structuring element followed by the horizontal structuring element. The imdilate function dilates the image.

Figure (4.9) dilated gradient mask

Step 4: Fill Interior Gaps

The dilated gradient mask shows the outline of the cell quite nicely, but there are still holes in the interior of the cell. To fill these holes we use the imfill function.The image will be as shown in figure(4.10).

Figure (4.10) binary image with filled holes

Step 5: Remove Connected Objects on Border

The cell of interest has been successfully segmented, but it is not the only object that has been found. Any objects that are connected to the border of the image can be removed using the imclearborder function. The connectivity in the imclearborder function was set to 4 to remove diagonal connections.As shown in figure(4.11).

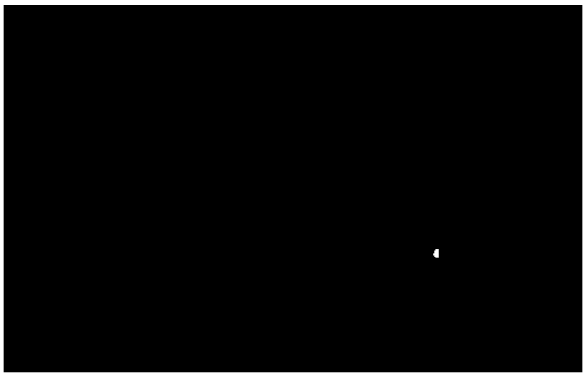


Figure (4.11) image cleared border image

Step 6: Smoothen the Object

Finally, in order to make the segmented object look natural, we smoothen the object by eroding the image twice with a diamond structuring element. We create the diamond structuring element using the strel function.An alternate method for displaying the segmented object would be to place an outline around the segmented cell. The outline is created by the bwperim function. As shown in figure(4.12).

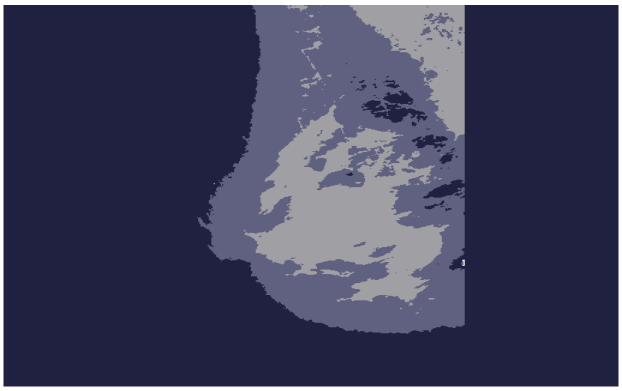


Figure (4.12) outlined original image

4.2.3 Feature extraction:

In this stage an extraction of features was selected to be applied as an input to the neural network. There are many types of feature extraction from digital mammogram:

***** Position feature.

*Shape feature.

*Texture feature, which it used here.

Texture features:

- The texture of a mammographic image is analyzed based on the difference between high and low gray levels in it.
- Texture is the ability to distinguish between normal and abnormal cases or patterns.
- It Provides information about textural characteristic of the image.
- And have 2 types:

➢ **First order (intensity based):** texture measure statistics calculated from an individual pixel and do not consider pixel neighbor relationship. Mean value, standard deviation was measured.

- **Mean Value:**

The mean gives the average intensity value of an image. Mammographic images that contain micro calcifications have a higher mean than those of normal images.

- **Standard Deviation:**

The standard deviation is a parameter closely associated with the mean. It refers to the dispersion of values in a mammographic image around the mean.

- **Second order (Gray Level co-occurrence matrix [GLCM]):**

texture can be characterized as the space distribution of gray scale levels in neighborhood so measures consider the relationship between neighbor pixels.

The Gray Level Co-occurrence Matrix (GLCM) texture measurement is a method to analyze image texture [5, 6]. It is a robust method that has been developed for calculating first and second order texture features from image. The GLCM matrix is a tabulation of how often different combinations of gray levels occur in an image. It considers the relationship between two pixels at a time, the reference and neighbor pixel. Usually, the Neighbor pixel is to the right of the reference pixel. Each cell of the matrix is normalized i.e., it Contains the probability of occurrence of a pixel pair and not just the count. GLCM indicate the possible pixel values of the image matrix [9]. A particular cell of the GLCM matrix refers to the number of occurrences of a pixel pair with the values specified by the corresponding row (i) and column (j). For example, the pixel pair (1, 1) occurs once in the image. Correspondingly, the first cell of the GLCM matrix holds the value 1.Similarly, the pixel pair (1,2) occurs twice and hence the second cell holds the value 2.Each of these cells is then normalized to obtain the probability of the occurrence of a pixel pair. Once the GLCM matrix has been formed, a set of formulae based on it helps us to calculate texture features [Appendix (A)].

Energy:

Energy represents the orderliness of a mammographic image. Energy is generally given by the mean squared value of a signal.

- **Entropy:**

Entropy measures the amount of disorder in a mammographic image. In the case of micro calcifications, the entropy value is high. This is because the variation in intensity values in the image is high due to the presence of white calcification spots.

- **Contrast:**

Contrast is a measure of the extent to which an object is distinguishable from its background. It represents the local variations present in an image, and calculates the intensity contrast between a pixel and its neighbor.

- **Cluster Shade:**

Cluster shade measures the uniformity of a mammographic image. When a large number of pixels of similar intensity exist together, homogeneity is high and is dependent on the gray scale value of the pixels that have similar intensities.

- **Sum of Square Variances:**

Sum of square variances is depends on the difference in gray level between adjacent pixels.

This feature puts relatively high weights on the elements that differ from the average value of P (i,j).

Both techniques of feature extraction shown in figure (4.13). and all extracted features calculations shown in Appendix (A).

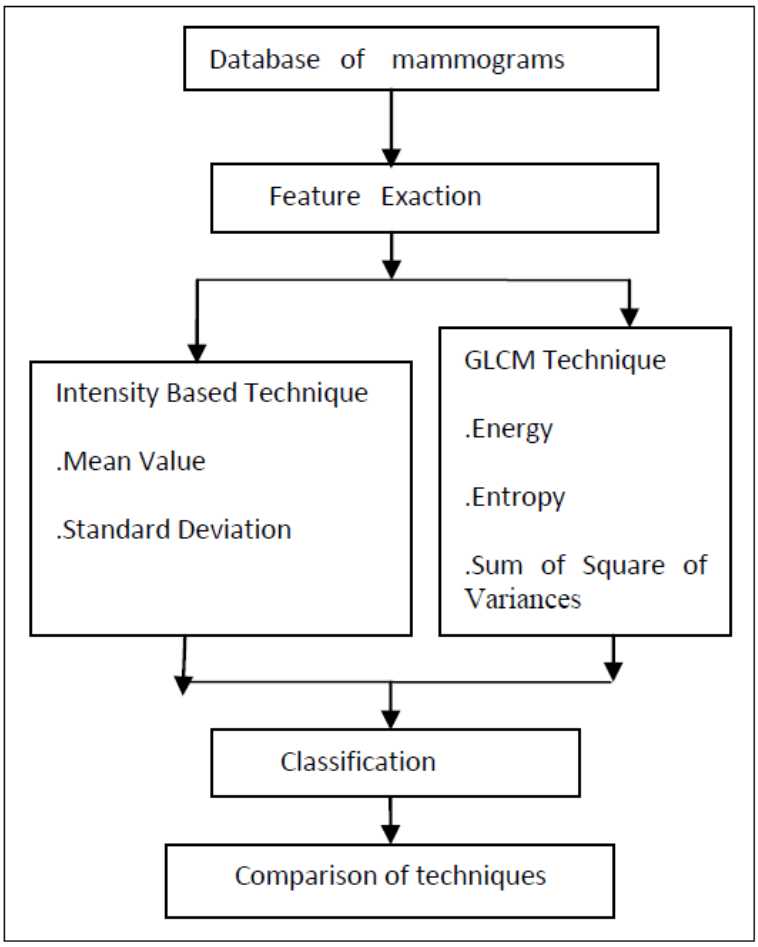


Figure (4.13) Feature extraction techniques

A list of the set of features calculated is explained here, AB= features of abnormal mammogram and NB= features of normal mammogram. As table (4.1) .

Table: 4.1- Features of GLCM and Intensity Based

FEATURE TYPE	ABI	NB1	AB2	NB2	AB3	NB3
ENTROPY	6.7281	6.1934	6.4167	6.9864	6.0263	6.5535
CONTRAST	4.79E+03	8.84E-01	1.00E+04	9.17E+04	1.74E+03	8.70E+03
CLUSTUR SHADE	0.4299	-0.0314	0.2847	-0.0394	0.1475	-0.0504
ENERGY	2.24E-05	1.51E-01	3.33E-05	1.65E-05	3.06E-05	1.75E-05
SUM OF SQUARE OF VARIENCE	0.0499	9.4709	0.0305	8.0352	0.032	6.0386
STANDARD DEVIATION	49.2184	13.361	71.4467	59.9481	60.4054	40.1557
MEAN	71.4791	56.5381	65.2625	125.5012	60.2067	103.8827

4.2.4 Classification:

For classification two types of neural networks was used (feed forward and back propagation neural networks) to classify the results of breast cancer.

The schematic representation of neural network with 'n' inputs, 'm'hidden units and one output unit. The extracted features are considered as input to the neural classifier. A neural network is a set of connected input/output units in which each connection has a weight associated with it. The neural network trained by adjusting the weights so as to be able to predict the correct class. The desired output was specified as 0 for non-cancerous and 1 for cancerous. The input features are normalized between 0 and 1. The classification process is divided into the training phase and the testing phase. During training, the features are extracted from the images in which the diagnosis is known. After training is over, the trained networks are stored to be used in the algorithm. Whenever an image is taken as input in the algorithm, it is simulated with the trained net-works and goes for testing the data. The accuracy, sensitivity, specificity of the classification is depends on the efficiency of the training [10, 13].

Matlab is a good programming toolbox package of version 2010a, provides functional software environment for creating neural network. The main goal of this package is to provide users with a set of integrated tools neural networks to create models of biological and simulate them easily, without the need of extensive coding.

Creation network:

The function and its own parameters below are used to create and define our neural network:

Net = new fit (inputs, targets);

The argument of this function is arranged as follows:

1. The inputs of neural network where it contains two important features based (intensity based, GLCM) extracted from the image.
2. Stated target for each stage performed by one dimensional binary array just one element of this array has value '1' and other elements are assigned to '0' that's to separate the desired target from other ones.

Training and Testing Stages:

The function and its own parameters below are used to train our neural network: net = train (net, inputs, targets); the function parameters are

1. Net: the neural network which created previously.
2. Inputs: inputs of the created neural network as defined before.
3. Targets: stated target the neural network.

4.2.5 Diagnosis:

The diagnosis will be done to compare between the two neural networks using measures for performance.

Measures for Performance:

A number of different measures are commonly used to evaluate the performance of the proposed method. These measures including classification, sensitivity, specificity, Mathews correlation coefficient (MCC) are calculated from confusion matrix using the equation 1, 2, 3&4. It returns a value from -1(inverse prediction) to +1(perfect prediction). The confusion matrix describes actual and predicted classes of the proposed method.

- **True Positive (TP) –** counts of all samples which are correctly called by the algorithm as being cancer.
- **False Positive (FP) –** counts of all samples which are incorrectly called by the algorithm as being cancer while they are normal.

- **True Negative (TN) –** counts of all samples which are correctly called by the algorithm as being normal.
- **False Negative (FN) –** count of all samples which are incorrectly called by the algorithm as being normal while they are cancer.

The performance of the classification algorithms was evaluated by computing the percentages of Sensitivity (SE), Specificity (SP), Accuracy (AC) and Mathews Correlation Coefficient (MCC), the respective calculations are as follows:

$$SE = TP/(TP+FN)*100 \quad \ldots (1)$$

$$SP = TN/(TN+TP)*100 \quad \ldots (2)$$

$$AC = (TP+TN)/(TN+TP+FN+FP)*100 \quad \ldots (3)$$

$$MCC = (TP \times TN - FP \times FN)/\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} \quad \ldots (4)$$

Chapter Five

Results and Discussions

5.1 The feed forward neural network results:

The effectiveness of the two texture feature extraction methods is trained and testing is done using neural classifier. The dataset used for this work is composed of each method, the architecture of the neural network; training and testing samples are same.

- In the experiment 1, Intensity based features are extracted and its classification is done using neural classifier.
- In the experiment 2, GLCM features are extracted and its classification is obtained.

The result shows that intensity based neural network is giving 94.11% classification rate and GLCM based neural network is giving 100% classification rate.

The confusion matrix for two different feature extraction methods was designed to detect the cancerous and non-cancerous where:

TP: True Positive. TN: True Negative.

FN: False Negative. FP: False Positive.

This is presented in Table (5.1) and (5.2).

Table (5.1) FFANN Intensity Based Confusion Matrix

Actual	Predicted	
	Cancerous	Non-Cancerous
Malignant	36(TP)	0(FP)
Benign	3(FN)	36(TN)

The Intensity Based Confusion Matrix results it was found that 36 dataset cancerous detected is (TP), and 3 dataset in benign is set as (FN) while 36 (TN) was detected as benign.

Table (5.2) FFANN GLCM Confusion Matrix

Actual	Predicted	
	Cancerous	Non-Cancerous
Malignant	36(TP)	0(FP)
Benign	0(FN)	39(TN)

The GLCM Confusion Matrix results it was found that 36 dataset cancerous detected is (TP), and 0 dataset in benign is set as (FN) while 39 (TN) was detected as benign.

The performance measures are calculated individually for the two different feature extraction methods are shown in Table (5.3). The proposed network was trained with all 75 tumor data set. These 75 cases are fed to the feed forward neural network with 7 input neurons, one hidden layer of 20 neurons and 4 output neurons. A Matlab software package version 2010 is used to implement the software in the current work. When the training process is completed for the training set (75 cases), the last weight of the network were saved to be ready for the testing procedure. The testing process is done for 75 cases. These 75 cases are fed to proposed network and their output is recorded for calculation of the sensitivity, specificity, accuracy and Mathews Correlation Coefficient of prediction. The two different features set has taken one is intensity based and another is GLCM based.

Table: (5.3) FFANN Evaluation Results

Feature Selection	No. of case	S E	S P	AC (%)	MCC
Intensity Based	75	0.88	01	94.11%	0.88
GLCM	75	01	01	100%	01

In order to evaluate the network 75 dataset was used in the system and the two extraction methods intensity based and GLCM was used, the table (5.3) represents the results of the two methods in their features.

5.2 The back Propagation neural network results:

- In the experiment 1, Intensity based features are extracted and its classification is done using neural classifier.
- In the experiment 2, GLCM features are extracted and its classification is obtained.

The result shows that intensity based neural network is giving 92.00% classification rate and GLCM based neural network is giving 100% classification rate. The confusion matrix for two different feature extraction method presented in Table (5.4) to Table (5.5).

Table (5.4) BPANN Intensity Based Confusion Matrix

Actual	Predicted	
	Cancerous	Non-Cancerous
Malignant	37(TP)	0(FP)
Benign	3(FN)	35(TN)

The Intensity Based Confusion Matrix results it was found that 37 dataset cancerous detected is (TP), and 3 dataset in benign is set as (FN) while 35 (TN) was detected as benign.

Table (5.5) BPANN GLCM Confusion Matrix

Actual	Predicted	
	Cancerous	Non-Cancerous
Malignant	40(TP)	0(FP)
Benign	5(FN)	30(TN)

The GLCM Confusion Matrix results it was found that 40 dataset cancerous detected is (TP), and 5 dataset in benign is set as (FN) while 30 (TN) was detected as benign.

The performance measures are calculated individually for the two different feature extraction methods are shown in Table (5.6). The proposed network was trained with all 75 tumor data set. These 75 cases are fed to the back propagation neural network with 7 input neurons, one hidden layer of 20 neurons and 4 output neurons. The 75 cases are fed to proposed network and their output is recorded for calculation of the sensitivity, specificity, accuracy and Mathews Correlation Coefficient, (MCC) of prediction.

Table: (5.6) BPANN Evaluation Results

Feature Selection	No. of case	S E	S P	AC (%)	MCC
Intensity Based	75	0.72	01	92.00%	0.69
GLCM	75	01	01	100%	01

In order to evaluate the network75 dataset was used in the system and the two extraction methods intensity based and GLCM was used, the table (5.6) represents the results of the two methods in their features.

5.3 Comparison between feed forward and back-propagation neural networks:

After a simulation, it was found that the back propagation neural network results is approximately the same with feed forward, but for the accuracy of detection the feed forward was better.

Table (5.8) and figure (5.1) show the comparison between the two neural networks from the sensitivity side. Where,

SE: the sensitivity of feed forward NN.

SE2:the sensitivity of back propagation NN.

Table (5.8) Comparison between two networks in SE

Feature Selection	No. of case	SE	S E2
Intensity Based	75	0.88	0.72
GLCM	75	01	01

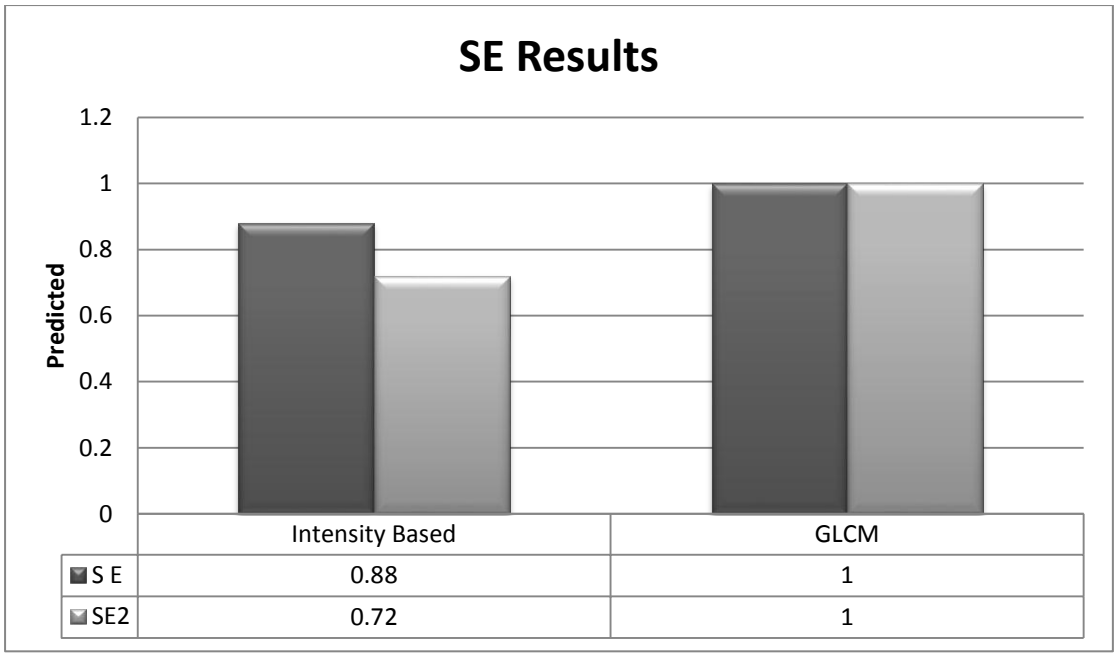


Figure (5.1) SE Results

Table (5.9) and figure (5.2) show the comparison between the two neural networks from the specifity side. Where,

SP: the specifity of the feed forward NN.

SP2: the specifity of the back propagation NN.

Table (5.10) Comparison between two networks in SP

Feature Selection	No. of case	S P	SP2
Intensity Based	75	01	01
GLCM	75	01	01

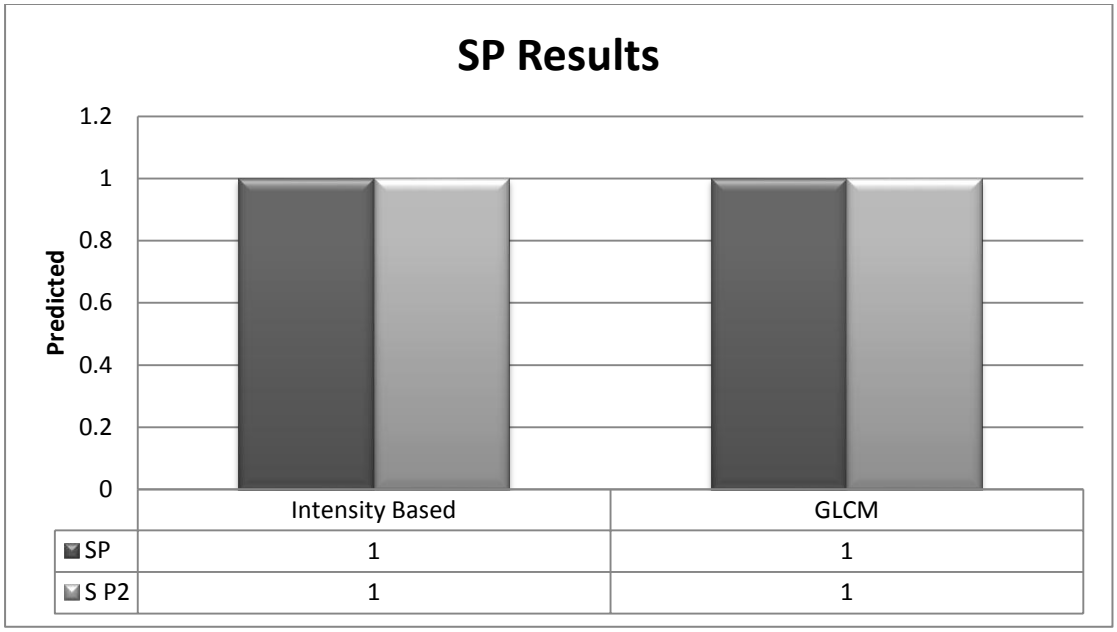


Figure (5.2) SP Results

Table (5.10) and figure (5.3) show the comparison between the two neural networks from the accuracy side. Where,

AC: the accuracy of the feed forward NN.

AC2: the accuracy of the back propagation NN.

Table (5.10) Comparison between two networks in AC

Feature Selection	No. of case	AC(%)	AC (%)2
Intensity Based	75	94.11%	92.00%
GLCM	75	100%	100%

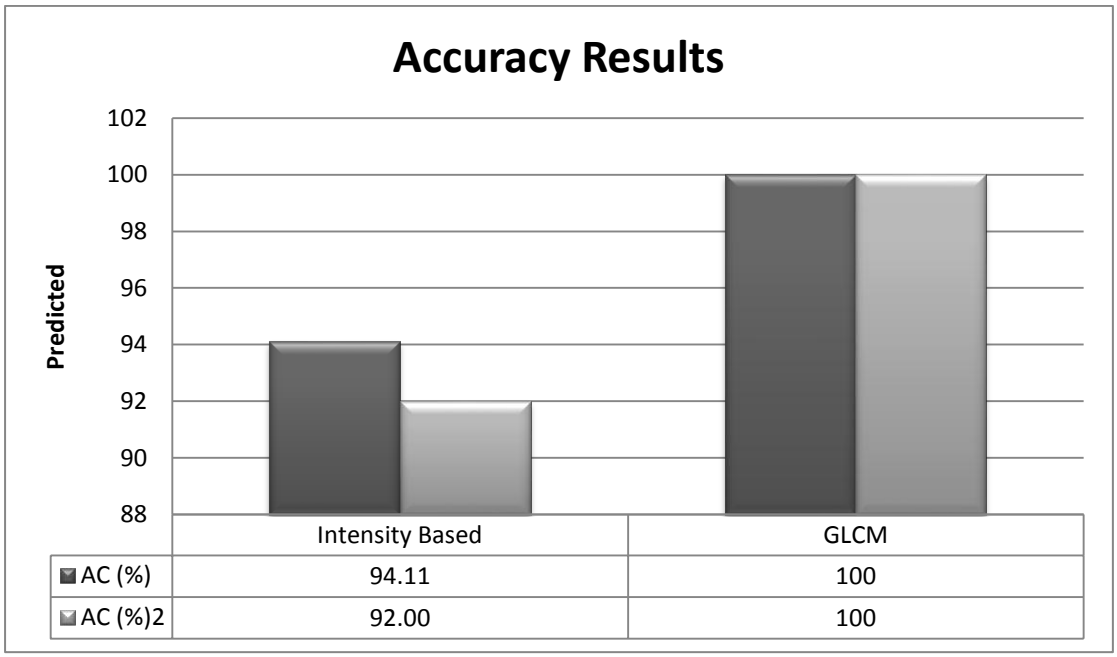


Figure (5.3) AC Results

Table (5.11) and figure (5.4) show the comparison between the two neural networks from the MCC side. Where,

MCC: the MCC of the feed forward NN.

MCC: the MCC of the back propagation NN.

Table (5.10) Comparison between two networks in MCC

Feature Selection	No. of case	MCC	MCC2
Intensity Based	75	0.88	0.72
GLCM	75	01	01

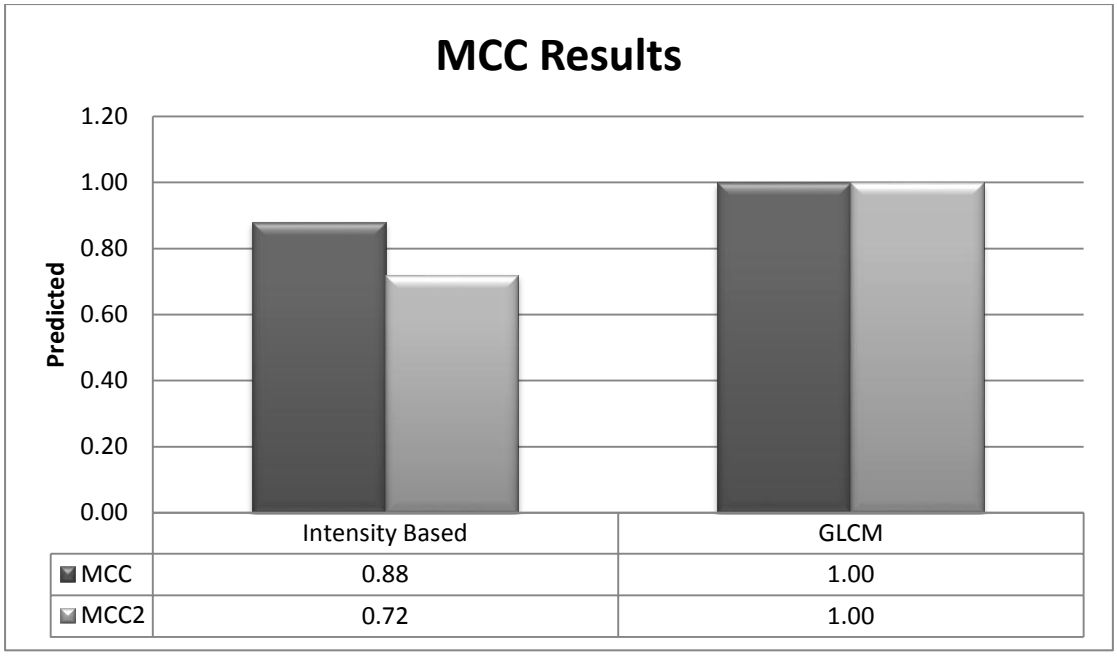


Figure (5.4) MCC Results

After testing the two extracting methods on two neural networks it was found that the accuracy of feed forward neural network is 94.11% accurate than the back propagation neural network which is 92.00%.

And there are no clear differences between the other performance measures.

Chapter Six

Conclusion and recommendations

6.1 Conclusion:

Breast cancer is one of the major causes of death among women. Early diagnosis through regular screening and timely treatment has been shown to prevent cancer. In this thesis an identification to the presence of breast cancer mass and calcification in mammograms and extracted clinically features and after this experiments we use ANN soft computing method for detect the cancer and easily differentiate the benign and malignant. This will help doctor to take or analyze in which stage of cancer the patient have and according to which he/she can take necessary and appropriate treatment steps.

This proposed method is low cost as it can be implemented in general computer .This paper is based on visual detection method of the processed mammogram images. In future aspect a real-time system can be implemented using suitable data acquisition software and hardware interface with digital mammography systems.

After testing the two extracting methods on two neural networks it was found that the accuracy of feed forward neural network is 94.11% accurate than the back propagation neural network which is 92.00%.

6.2 Recommendations:

This thesis was based on visual detection method of the processed mammogram images. In future aspect

- A real-time system can be implemented.
- using suitable data acquisition software and hardware interface with digital mammography systems.
- To apply other neural network types.

References

1. McCann, executive publisher, Judith A. Shilling (2008). Nursing-Ambler, PA: Lippincott Williams & Wilkins. p. 99. ISBN 978-1-58255-668-0.

2. Santoro, E., DeSoto, M., and Hong Lee, J (February 2009). "Hormone Therapy and Menopause". National Research Center for Women & Families.

3. Light Pollution as new risk factor for human Breast and Prostate Cancers- Haim,Abraham; Portnov, Biris P. (2013),ISBN 978-94-007-6220-6

4. Boris Pasche (2010). Cancer Genetics (Cancer Treatment and Research). Berlin: Springer. pp. 19–20. ISBN 1-4419-6032-5.

5. Kumar, Vinay; Abul Abbas (2010). Robbins and Cotran Pathologic Basis of Disease. Philidelphia: Saunders, an imprint of Elsevier inc. p. 1090. ISBN 978-1-4160-3121-5.

6. Olson, James Stuart (2002). Bathsheba's breast: women, cancer & history. Baltimore: The Johns Hopkins University Press. pp. 9–13. ISBN 0-8018-6936-6.

7. Yalom, Marilyn (1997). A history of the breast. New York: Alfred A. Knopf. p. 234. ISBN 0-679-43459-3.

8. Aronowitz, Robert A. (2007). Unnatural history: breast cancer and American society. Cambridge, UK: Cambridge University Press. pp. 22–24. ISBN 0-521-82249-1.

9. Alfredo Morabia (2004). A History of Epidemiologic Methods and Concepts. Boston: Birkhauser. pp. 301–302. ISBN 3-7643-6818-7. Retrieved 2007-12-31.

10. Sulik, Gayle A. (2010). Pink Ribbon Blues: How Breast Cancer Culture Undermines Women's Health. USA: Oxford University Press. pp. 200–3. ISBN 0-19-974045-3.

11. Connie Henke Yarbro, Debra Wujcik, Barbara Holmes Gobel, ed. (2011). Cancer nursing: principles and practice (7 ed.). Jones & Bartlett Publishers. pp. 901–905. ISBN 978-1-4496-1829-2.

12. http://www.cancer.ca/en/cancer-information/cancer-type/breast/anatomy-and-physiology/?region=on.

Appendix (A)

Features Calculations

1. Mean calculated from the image as per the following equation.

$$\mu = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} P(i,j)$$

Where '*i' i*ndicates the rows of the image, 'j' indicates the columns of the image and P (i, j) is the cell denoted by the row and the column of the image.

2. Standard deviation is given as:

$$SD = \sqrt{(mean)^2}$$

3. Energy calculated from the image as per the following equation.

$$E = \sum_{i,j=0}^{n-1} P(i,j)^2$$

Where '*i' indicates* the rows of the GLCM matrix, indicates the columns of the GLCM matrix and P (i, j) is the cell denoted by the row and the column of the GLCM matrix.

4.Entropy calculated from the image as per the following equation.

$$H = -\sum_{i=0}^{n} P(i,j) \log P(i,j)$$

5.Contrast calculated from the image as per the following equation.

$$C = \sum_{i,j=0}^{n-1} (i.j)^2 P(i,j)$$

Where, n denotes the number of pixels in the image and P (i, j) is the cell denoted by the row and column of the image.

6. Cluster shade is calculated from the image as per the following equation.

$$C_s = -\sum_{i,j=0}^{n-1} P_{(i,j)} (log_2 P_{i,j})$$

Where, P (i, j) is the cell denoted by the row and column of the image.

7. Sum of square variances is calculated from the image as per the following equation.

$$S_v = \sum_{i,j=0}^{n-1} P_{(i,j)} (i - j)^2$$

Appendix (B)

Segmentation Code:

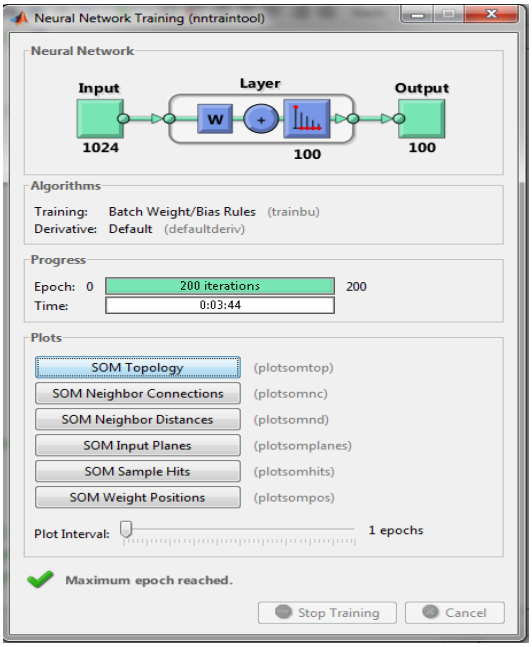
```
clear all;
clc;
close all;
%**************************1*****************
figure(1);
rawData1 = importdata('mdb001.pgm');
rawData56=importdata('mdb056.pgm');
imshow(rawData1)
figure(2);
A = imnoise(rawData1,'salt & pepper',0.02);
A1 = imnoise(rawData56,'salt & pepper',0.02);
RImg=medfilt2(A);
RImg1=medfilt2(A1);
imshow(RImg);
figure(4);
tt = adpmedian(RImg,3);
tt1 = adpmedian(RImg1,3);
imshow(tt);
figure(5);
AA = adapthisteq(tt,'clipLimit',0.02,'Distribution','rayleigh');
AA1 = adapthisteq(tt1,'clipLimit',0.02,'Distribution','rayleigh');
imshow(AA);
figure(6);
[clusters, result_image, clusterized_image] = kmeansclustering(AA, 3);
[clusters, result_image1, clusterized_image] = kmeansclustering(AA1, 3);
imshow(result_image);
```

```
figure(7);
M1 = mean( result_image);
M2 = mean( result_image1);
mm=M1-M2;
plot(mm);
mm
figure(22);
M3 = median( result_image);
M4 = median( result_image1);
mmm=M3-M4;
plot(mmm);
mmm
%*********************************
figure(8);
 [mm fval] = ga(@rastriginsfcn, 2)
 plot (mm);
 figure(20);
 [mmm fval] = ga(@rastriginsfcn, 2)
 plot (mmm);
[junk threshold] = edge(result_image, 'sobel');
fudgeFactor = .5;
BWs = edge(result_image,'sobel', threshold * fudgeFactor);
figure, imshow(BWs), title('binary gradient mask');
se90 = strel('line', 3, 90);
se0 = strel('line', 3, 0);
BWsdil = imdilate(BWs, [se90 se0]);
figure, imshow(BWsdil), title('dilated gradient mask');
BWdfill = imfill(BWsdil, 'holes');
figure, imshow(BWdfill);
```

```
title('binary image with filled holes');
BWnobord = imclearborder(BWdfill, 4);
figure, imshow(BWnobord), title('cleared border image');
seD = strel('diamond',1);
BWfinal = imerode(BWnobord,seD);
BWfinal = imerode(BWfinal,seD);
figure, imshow(BWfinal), title('segmented image');
BWoutline = bwperim(BWfinal);
Segout = result_image;
Segout(BWoutline) = 255;
figure, imshow(Segout), title('outlined original image');
```

Appendix(C):

Neural Network Creation



creating neural network [Matlab wizard]

When using the neural network the parameters that should be used is result_image as an input to the network with a creation diminution of 10 by 10 matrixes

```
inputs = result_image;
dimension1 = 10;
dimension2 = 10;
```

to create net map

```
net = selforgmap([dimension1 dimension2]);
```

to start training the network

```
% Train the Network
[net,tr] = train(net, inputs);
```

Testing the network depending on training

```
% Test the Network
outputs = net(inputs);
```

view network result

```
% View the Network
View (net)
```