# BREAST CANCER DETECTION USING COMPUTATIONAL INTELLIGENCE

By

SITI AISHAH BINTI FADILLULLAH

PROJECT DISSERTATION

Submitted to the Electrical & Electronics Engineering Program
in Partial Fulfillment of the Requirements
for the Degree
Bachelor of Engineering (Hons)
(Electrical & Electronics Engineering)

Universiti Teknologi Petronas
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

# CERTIFICATION OF APPROVAL

**BREAST CANCER DETECTION USING COMPUTATIONAL INTELLIGENCE**

by

Siti Aishah Fadillullah

A project dissertation submitted to the
Electrical & Electronics Engineering Program
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
Bachelor of Engineering (Hons)
(Electrical & Electronics Engineering)

Approved:

Ms Zazilah May
Project Supervisor

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK

June 2005

# CERTIFICATION OF ORIGINALITY

This is to certify that the author is responsible for the work submitted in this project, that the original work is the author's except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

Siti Aishah Fadillullah

# ABSTRACT

Mammograms are the best tool to detect an early disease of breast cancer. In mammography, medical experts look for clustered microcalcifications and irregular density masses. As microcalcification is a tiny speck of calcium in breast, it appears as white spot in mammogram. Problem occurred when the clinician reads the mammograms using a magnifying glass, as it is difficult to detect calcification because there is a wide range of abnormalities and it also due to the small size and their similarity with other tissue structure. One of the problems is to distinguish between malignant and benign tumors. Thus, the objectives of this project are to enhance mammogram image using image processing technique and to provide a pattern recognition system by signifying whether further investigation is needed, therefore it may assist medical expert in detection of breast cancer. Accordingly, the scope of this project is based on the pattern recognition system, which includes preprocessing, feature extraction, and classification. The task for the project is divided into two parts. The first part is the enhancement of the image and the detection of calcification. The second part of the project is to design, develop, and test the network whether it run as expected. As the result, mammogram images have been processed through image processing by using MATLAB, and opening morphological operation has been used for the detection. A pattern recognition system has been developed by the use of neural network. As a conclusion, a successful implementation of pattern recognition system as one way to detect breast cancer could help medical field in diagnosing breast cancer.

# ACKNOWLEDGEMENTS

In the name of God, the Compassionate, the Merciful. The author would like to thank Him for the strength, skill, knowledge, patience and health that have been given through this time.

It is pleasure for the author to acknowledge the effort and assistance that have been shown by many people whose name is not mention on the cover. Without them, accomplishment of this report would have been impossible.

First and foremost, a million of thanks to the most supportive and encouraging supervisor, Ms Zazilah May, who had given all the knowledge and guidance in accomplishing this final year project. Despite her busy schedule, there always be an allocate time for her to discuss any arising matters.

The author also would like to give her greatest gratitude to Ms Kavitha for her assistance and dedication towards the accomplishment of the project. Without her expertise and skills, the project may not be able to complete within the given time frame.

Thank you to Dr. Seelaya, Dr. Shafie and stuff of Hospital Besar Ipoh, Perak for their generous information and source for the project.

Last but not least, the author would like to thank all the colleagues that have contributed in accomplishing this project. Thank you very much for all the supports and constructive comments.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ROI | Region of Interest |
| ROC | Receiver Operating Characteristics |
| TIFF | Tagged Image File Format |
| NN | Neural Network |
| RMSE | Root Means Square Error |
| No. | Number |
| R | right |
| L | left |

# CHAPTER 1
# INTRODUCTION

## 1.1 Background of study

Cancer begins in cells that behave abnormally and they are called a growth or tumor. However not all tumors are cancer, because they can be benign or malignant. Breast cancer is a malignant tumor that begins in the tissues of the breast [1], and there are several types of breast cancer such as ductal carcinoma and lobular carcinoma.

In Malaysia, breast cancer is one of the most common causes of death in women. The incidence of breast cancer in Malaysia is estimated to be around 27 per 100,000 populations, with close to 3,000 new cases annually [2].

Mammograms are the best tools to detect the early disease of breast cancer. In mammograms, usually doctor or radiologist will looks for clustered microcalcifications and irregular density masses. As microcalcification is a tiny speck of calcium in breast, it appears as white spot in mammogram. It may appear alone or in clusters. The white spot do not always mean cancer is present, it may be microcalcification, or it might also be an artifact.

The project will covered the process of enhancing the image, extracting the features, recognizing the pattern, and classifying them in order to assist medical experts in interpreting the mammograms.

## 1.2 Problem statement

Mammography is the most efficient way to detect early sign of breast cancer. However reading mammograms is difficult because there is a wide range of abnormalities and it also due to the small size and their similarity with other tissue

structure. One of the problems is to distinguish between malignant and benign tumors.

According to Susan Orel quoted in [3], the biggest misconception about mammography is that it picks up every breast cancer, and in fact, mammography misses at least 10 percent of breast cancer. Usually the abnormality in breast that seems to be a cancer but turn out to be normal is called false positive, and the unidentifiable breast cancer in a mammography is called false negative.

This project can help with the detection of breast cancer so that the suspicious area of abnormality can be recognized. The study on image enhancement using image processing technique, and pattern recognition using computational intelligence, which is neural network may assist medical experts to examine the presence of breast cancer.

## 1.3 Objectives

- To enhance mammogram image through image processing technique using MATLAB.
- To provide a pattern recognition system using neural network.
- To help medical expert in detecting and determining various stage and locations of breast abnormalities, and to check whether further investigation is needed.

## 1.4 Scope of study

The scope of this project is based on the pattern recognition system, which includes image processing, feature extraction, and classification. Understanding the terms and technique in medical imaging is desirable.

The project basically focused on mammogram's image, the enhancement of the image, techniques used in processing the image, and other related topics. Preprocessing stage will covered image enhancement such as filtering which is the essential process in order to extract the features.

MATLAB is used to process the image, and for the computational intelligence technique, neural network model will be applied. And the output produced can be used to assist doctor in detecting breast cancer.

# CHAPTER 2
# LITERATURE REVIEW

## 2.1 Breast cancer

Breast is a gland that can produce milk, and each breast sits on chest muscles that cover the ribs [4]. It is made up mainly of lobules which are milk-producing glands, ducts which are the milk route that connect the lobules to the nipple, and stroma which is the fatty tissue and connective tissue surrounding the ducts and lobules, blood vessels, and lymphatic vessels. Figure 1 shows the normal breast structure. Most breast cancers begin in the ducts (ductal), some in the lobules (lobular), and the rest in other tissues [5].

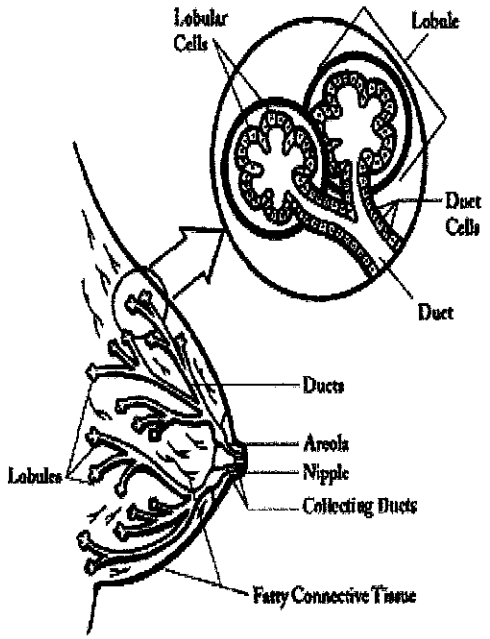



Figure 1 Normal breast structure

Uncontrolled cells in breast that produce extra cell can form mass of tissue called tumor. There are two types of tumor: benign tumor and malignant tumor. Benign tumor is not cancer and it is rarely life threatening because the cell does not spread to tissues around them. But malignant tumor is cancer. Usually malignant tumor is more serious and may be life-threatening as the cell can invade and damage nearby tissue and organs

Many medical experts categorize a cancer according to an established breast cancer staging system based on the size of the tumor, the extent to which the tumor is involved with the skin, muscles, and other tissues next to it, and lymph node involvement [3].

Breast cancer can be divided into seven stages; stage 0, stage I, stage IIA and IIB, stage IIIA and IIIB, stage IV (for more details see APPENDIX I) [1]. The stages reflect the seriousness of the case.

## 2.2 Mammograms

Mammogram is an x-ray examination of breast (see Figure 2). A screening mammogram is an x-ray examination of the breast in a woman who has no breast complaints (asymptomatic). And a diagnostic mammogram is an x-ray examination of the breast in a woman who either has a breast complaint (for example, a breast mass, nipple discharge, etc.) or has had an abnormality found during a screening mammogram.

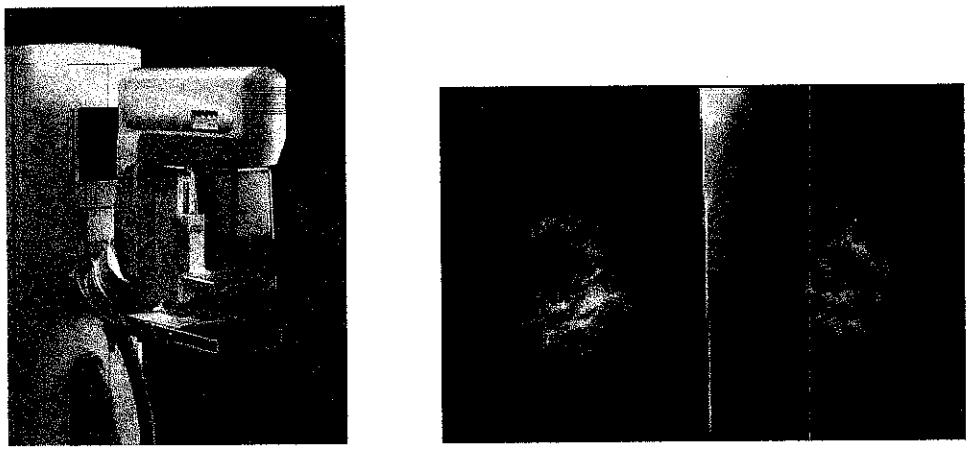


Figure 2 Mammogram machine (left) and side-to-side mammogram image of both breast (right)

Conventional mammography creates the image of breast tissue on film. Because mammograms are not perfect, there is a need for new technologies that are better and able to detect breast cancer. With digital mammography it is possible to capture and display the x-ray information on computers, without the use of film [6]. It is then possible to enhance the quality of the image and even magnify the view of specific areas of the breast. But either conventional or digital, both mammographies are to look for abnormalities in breast such as calcifications, which are tiny mineral deposits within the breast tissue that looks like white small spots on the films.

There are two types of calcifications: macrocalcifications and microcalcifications. Macrocalcifications are coarse calcium deposit that are related to non-cancerous conditions and do not require a biopsy. Microcalcifications are tiny specks of calcium in the breast. They may appear alone or in cluster. The presence of microcalcification does not always mean cancer is present. Usually radiologists judge the presence of cancer by looking at the characteristics of the calcifications. Detail list of characteristic for calcification can be obtained in [15]. Another abnormality that may appear in a mammogram is mass. It may occur with or without calcification. Mass is either cyst which is benign collection of fluid in the breast, or maybe cancer (depends on size, shape, and margin).

Apart from digital mammography, which image can be enhance directly, nowadays computer-aided detection and diagnosis can be use to enhance the conventional mammography's image by digitizing the image. The M1000 Image Checker is one such device that has been approved by the US Food and Drug Administration (FDA) for use in reviewing mammograms [5]. This device can detect some cancer that the doctor might miss. In order to clarify the effectiveness of the computer-aided detection and diagnosis more technical refinements and studies that help to clarify their role in breast cancer detection is needed.

## 2.3 Image processing

Image is a two-dimensional function, $f(x, y)$ where $x$ and $y$ are spatial (plane) coordinates [14]. The amplitude of f at any pair of coordinates $(x, y)$ is called the intensity of grey level of the image at that point. In digital image, the $x$, $y$, and

amplitude values of $f$ are all finite and discrete quantities. Processing digital image by means of digital computer is called digital image processing.

Basically, digital image processing involves a computer to process images and two pieces of special input/output equipment: an image digitizer and an image display device. In processing an image there are many steps that can be applied such as image formation (image acquisition), image restoration, image enhancement, image analysis, image reconstruction, and compression.

### *2.3.1 Digital image formation*

Digital image formation is a process of capturing the image. The system basically consists of image acquisition, and digitizer. Image acquisition is done to generate digital image from sensed data, which includes optical system and sensor. An analog signal is transformed to digital by a digitizer.

In order to convert analog signal to digital form, we need to sample the function in both coordinates and amplitude. Digitizing coordinate values is called sampling, and digitizing amplitude values is called quantizing. Each digital image formation subsystem introduces a deformation or degradation to the digital image, such as geometrical distortion, noise, and nonlinear transformation.

### *2.3.2 Digital image enhancement*

Enhancement techniques is done to bring out detail and to highlight certain features of interest in an image. In another words, it is to improve the quality of image (to look better) in terms of contrast, image sharpening, noise reduction, and so on. Image enhancement is not to increase the inherent information content in data, but it is to emphasize certain specified image characteristics by increase dynamic range of chosen features so they can detect easily.

Image enhancement techniques can be classified into two methods: spatial domain and frequency domain. Spatial domain methods are based on direct

manipulation of gray values of pixels in an image. Frequency domain methods are based on modifying the Fourier transform of an image.

Image enhancement includes grey level and contrast manipulation, noise reduction, edge crispening and sharpening, filtering, interpolation and magnification, pseudocoloring and so on.

- Noise reduction

  Noise can be introduced into an image. It depends on how the image is created. There is different ways to remove or reduce noise in an image, as different methods are better for different kinds of noise. The methods available include linear filtering, median filtering, and adaptive filtering. As an example, best-suited filter for salt and pepper noise is median filter. Figure 3, Figure 4 and 5 shows the effect of each filter on salt and pepper noise.

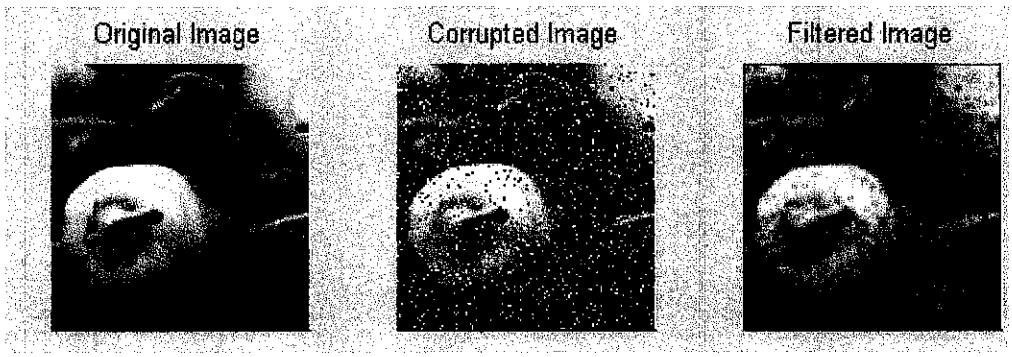



Figure 3 Effect of averaging filter on salt and pepper noise


Figure 4 Effect of median filter on salt and pepper noise

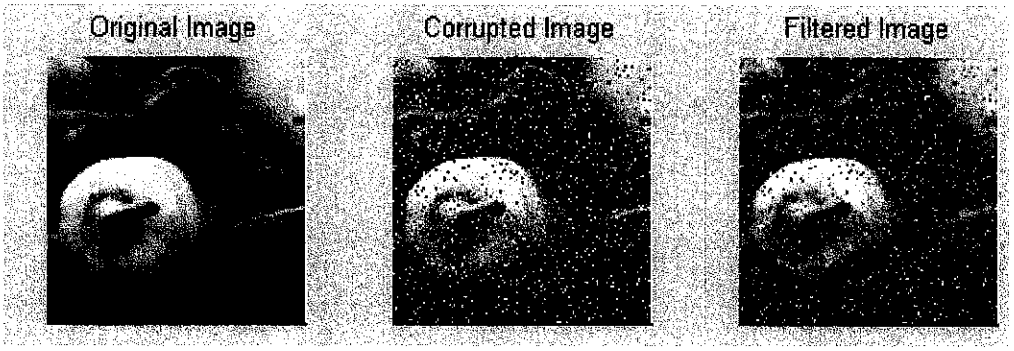



Figure 5 Effect of adaptive filter on salt and pepper noise

- Filtering technique

Filter is to remove noise or to enhance edge and small details in an image. Lowpass filter is used to smooth the image and it is used for noise removal. Lowpass filter can blur the image as it suppressed high-frequency coefficient and enhanced the low-frequency coefficient. Gaussian lowpass filter yields a lowpass filter with smooth behavior in both domain [14].

When the low-frequency coefficient is suppressed and high-frequency is boosted, it is called a highpass filter. Highpass filter is possible to sharpen image as the edge and small details correspond to high-frequency coefficient.

During averaging and lowpass filtering, each pixel is replaced by the weighted average of its neighborhood pixels [17], that is

$$v(m,n) = \sum_{(k,l)\in W}\sum a(k,l)y(m-k, n-l) \qquad (1)$$

Where $v(m,n)$ is the input and an output image, $W$ is a suitable chosen window, and $a(k, l)$ is the filters weight.

Median filter is an order-statistics filter [12]. It replaces the value of a pixel by the median of the grey levels in the neighborhood of the pixel, and it is better in reducing random noise without reducing the sharpness of the image. The effect of median filter on salt and pepper noise is as in Figure 4.

- Contrast enhancement

  Contrast enhancement or contrast stretching is a point operation that is used to expand the contrast of the features of interest so that they occupy a larger portion of the displayed grey-level range [12]. It is to increase dynamic range of the gray levels in the processed image. Figure 6 shows the effect of contrast enhancement on an intensity image.

Figure 6 Original image (left), and the output of contrast enhancement

### *2.3.3 Image analysis*

Image analysis is related to make quantitative measurement from an image to produce a description of it. It requires extraction of certain features that aid in the identification of the object. Image analysis consists of edge and line detection, texture analysis, segmentation, region-of-interest (ROI) processing, feature measurement, and so on.

Segmentation is one of the most important steps to analyze image data. Its main goal is to divide an image into parts that have strong correlation with objects or areas of the real world contained in the image [21]. Gray-level thresholding is the simplest segmentation process and it is computationally inexpensive and fast.

Features extraction is to reduce data by measuring certain "properties" that distinguish input pattern. There are many techniques and approach for feature extraction. There is Fourier transform domain feature extraction, Walsh-Hadamard transform (WHT) domain feature extraction, invariant feature extraction, and texture features.

Pattern recognition is one of the aspects in analyzing an image. Statistical pattern recognition assumes that the image may contain one or more objects and that each object belongs to one of several predetermined types, categories, or pattern classes [12]. There are three major phases in pattern recognition: image segmentation, feature extraction, and classification. Pattern recognition systems usually consider a feature space onto which the observation vector is first mapped. The feature vector is then used to decide the class to which the observation vector belong base on the measured objects.

Classification can be described as the process of mapping a feature vector from feature space to class membership space. Conventional methods include statistical and syntactic techniques. In the statistical approach, a set of features is extracted from the input pattern, and partitioning the features space carries out the classification. One way of pattern recognition techniques is to group them into supervised and unsupervised methods.

## 2.4 Feature extraction

Feature extraction is a process where input variables (vectors) are selected for the design of a neural network especially in a pattern recognition decision aid [10]. According to the authors, type of variable to be used in neurons of the input layer must be first verified as these variables are useful in distinguishing between two classes.

Isaac N. Bankman et al [18] has presented a segmentation algorithm and compare it to the multitolerance region growing algorithm of Shen et al and active contours. The segmentation algorithm operates without threshold or window selection or parametric data models, which is called hill climbing.

The author has stated that Shen at al have done automatic thresholding that uses a growth tolerance parameter that changes in a small range with a step size that depends on the seed pixel. Three features are extracted from each region grown with different tolerance level: shape compactness, centre of gravity, and size.

Isaac N. Bankman et al, also stated that the width of the smallest microcalcification consider in his study was about 0.25mm and the majority of the microcalcifications are in the range of width of 0.3 to 0.5mm. The author used a circle of 0.2mm diameter around the local maximum pixel as the initial position of the active contour (24 8-connected pixels). By segmentation algorithm in [18], they had extract four features: contrast, relative contrast, area, and edge sharpness.

According to A. Wróblewska et al [20], the first step in automatic feature selection method is an extraction of a broad feature set, containing promising features found in many publications, and this large set will be reduced in order to find features essential and valuable for classification of microcalcifications. The authors have divided all evaluated features into three groups, which are texture features, shape features, and scalar features.

D Betal et al [21], applied mathematical morphology algorithm to describe microcalcification shape in terms of the presence or absence of infoldings, elongation, narrow irregularities and wide irregularities. An ROC analysis was performed to investigate the effect on sensitivity and specificity of the proportion of the nine neighbors that agreed with the true calcification.

Masses can be distinguished by shape, size, and margin characteristics. And calcifications can be characterized by size, number, morphology, distribution, and heterogeneity. Figure 7 shows mass shape and margin characteristics.

According to "Interactive Mammography Analysis Web Tutorial" [22], masses are three-dimensional lesions that may represent a localizing sign of breast cancer. They are described by their location, size, shape, margin characteristics, x-ray attenuation (radio density), effect on surrounding tissue, and any other associated findings (i.e. architectural distortion, associated calcifications, skin changes). Depending on the morphologic criteria of the mass, the likelihood of malignancy can be established.

Aside from masses, a suspicious single geographic abnormality can also be classified by calcifications. Calcifications are analyzed according to their size, shape, number, and distribution. The general rule is that larger, round or oval shaped calcifications uniform in size has a higher probability of being associated with a benign process. And smaller, irregular, polymorphic, branching calcifications heterogeneous in size and morphology are more often associated with a malignant process.

Number of calcification that made up a cluster has been used as an indicator of benign and malignancy. While the actual number itself is arbitrary, radiologists tend to agree that the minimum number of calcifications be four, five, or six to be of significance. Any number of calcifications less than four will rarely lead to the detection of breast cancer in and of itself. Again, as with all criteria in mammography analysis, no number is absolute and two or three calcifications may merit greater suspicion if they exhibit worrisome morphologies.

Area is computed as the number of pixels in the grown region. It is measured by counting the number of pixels inside and including the boundaries. It is relates to the size of calcification. Most radiologists place calcifications 0.5 mm or less to have a high probability of association with cancer; and calcifications of 2.0 mm or larger are typical of a benign process. The smallest visible calcification on a mammogram is approximately 0.2 - 0.3 mm.

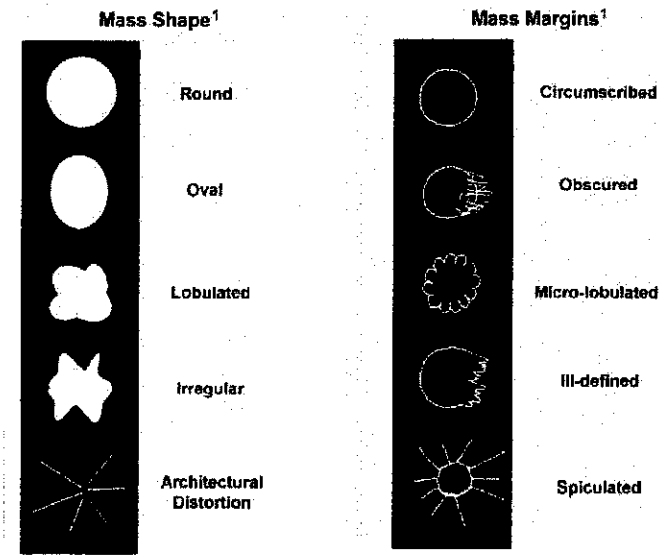



Figure 7 Mass shape and margins characteristics ( Diagram adapted from BB Kopans Breast Imaging (J.B. Lippincott Co., Philadelphia; 1989)

Perimeter measurement is to measure an object's perimeter to establish that the boundary of an object is polygon having a vertex at the center of each boundary pixel. Perimeter can also be measured by summing center-to-center distance between adjacent pixels on the boundary. The perimeter is measured after the *bwperim* process, which is after perimeter determination by applying *sum([data2.Area])* to the *bwperim* image. Thus to calculate the total perimeter, total area of *bwperim* is divided with the number of calcifications;

$$Mean\ Perimeter = sum([data2.Area]) / number\ of\ objects \qquad (2)$$

Eccentricity and orientation is a scalar vector. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1, which is when approaching 0 represents a circle and approaching 1 represents a line segment. Orientation is the angle (in degrees) between the x-axis and the major axis of the ellipse that has the same second-moments as the region. Solidity is also a scalar vector. It is the proportion of the pixels in the convex hull that are also in the region. Solidity is computed as *Area/ConvexArea.*

Mathematically, area of a circle is calculated as $pi*r^2$ while the perimeter is calculated as $2*pi*r$. By computing the '*equivDiameter*' the diameter of an object can be obtain. Thus circularity can be calculated as below:

$$Circularity = (4*pi*area)/perimeter^2 \qquad (3)$$

## 2.5 Neural network

Neural network operates in parallel and it is inspired by biological nervous systems. The network can be train to perform particular function by adjusting the values of the connection (weight) between elements. This is to get a specific target output.

According to A. Wróblewska et al, the number of input layers neurons was the same as a size of feature vector, and hidden layer neurons was experimentally set according to the number of recognized classes. In a single-input neuron (as in Figure 8), a scalar input $p$ is multiplied by the scalar weight $w$ to form $wp$. The bias $b$ has a constant input of 1. Transfer function net input $n$ is the sum of the

weighted input $wp$ and the bias $b$. The net input $n$ goes to transfer function $f$, which produces the scalar neuron output $a$. [19]

Thus the neuron output is calculated as

$$a = f(wp + b) \qquad (4)$$

In order to satisfy some of the problem that the neuron attempt to solve, a transfer function needs to be chosen. Hard limit transfer function take argument value between 0 and 1 and mostly used for decision making. Linear transfer function used as linear approximators. The sigmoid transfer function *logsig* takes the input of any finite value and gives the output into the range of 0 and 1.

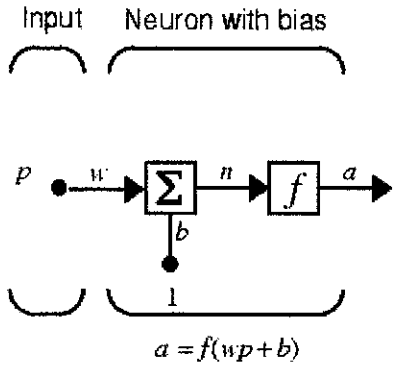



Figure 8 Single-input neuron

According to MATLAB Neural Network Toolbox [23], backpropagation was created by generalizing Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer function. Input vector and corresponding target vector are used to train until an approximation of the function, which relate the input and the output is generated. Once the network is trained, the network is able to approximate a set of inputs to certain accuracy without providing output. Multilayer feedfoward network is most commonly used network architecture for the backpropagation algorithm. Multilayer network often use the log-sigmoid transfer function *logsig* (as in Figure 9). And occasionally, the linear transfer function *purelin* (Figure 9) is use in backpropagation networks.

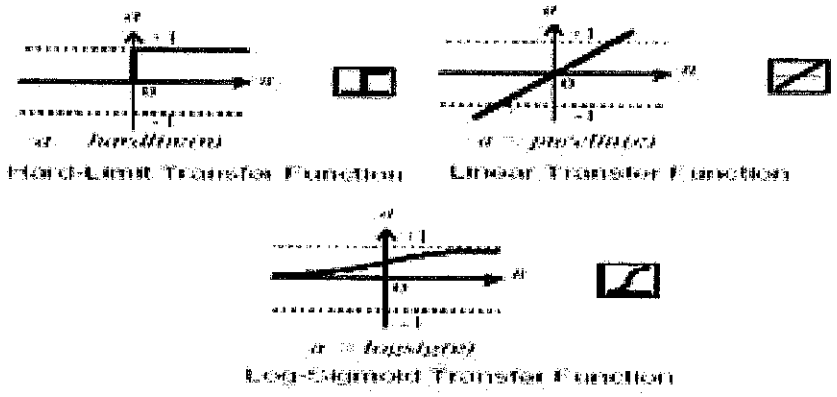


Figure 9 Transfer function

Basic backpropagation network architecture is shown in Figure 10. The number of hidden layers in a Feed Forward network is often one or more layers. According to Neural Network Toolbox, there are no rules leading the amount of layers and number of neurons. Normally trial and error approach is used to determine the best construction of network that can be specified before the network is trained. There are generally four steps in training process:

1. assemble the training data
2. create network object
3. train network
4. simulate the network response to the new input

Neural network is trained to classify the pattern of calcifications. The training process requires a set of inputs and its targets. Weights and biases are iteratively adjusted to minimize the network performance function (adjust to get the minimum error). In backpropagation, weights are moved in the direction of negative gradient.

According to MATLAB Neural Network Toolbox, gradient descent algorithm can be implemented by incremental mode and batch mode. Examples of batch mode are such as batch training (*train*), batch gradient descent (*traingd*) and batch gradient descent with momentum (*traingdm*).

The training algorithm *traingd* and *traingdm* are often too slow for practical problems. Fast algorithm can be generalized as those that use heuristic techniques and those that use standard numerical optimization techniques.

Heuristic is based on the analysis of the performance of the standard steepest descent algorithm. Examples of heuristic training algorithm are variable learning rate backpropagation (*traingda*) and resilient backpropagation (*trainrp*).

And example of algorithm that uses the standard numerical optimization techniques is conjugate gradient (*traincgf, traincgp, traincgb, trainscg*), Quasi-Newton (*trainbfg, trainoss*), and Levenberg-Marquardt (*trainlm*). The suitable training algorithm for pattern recognition network is resilient backpropagation (*trainrp*), conjugate gradient algorithms (*trainscg*), and Levenberg-Marquardt (*trainlm*).

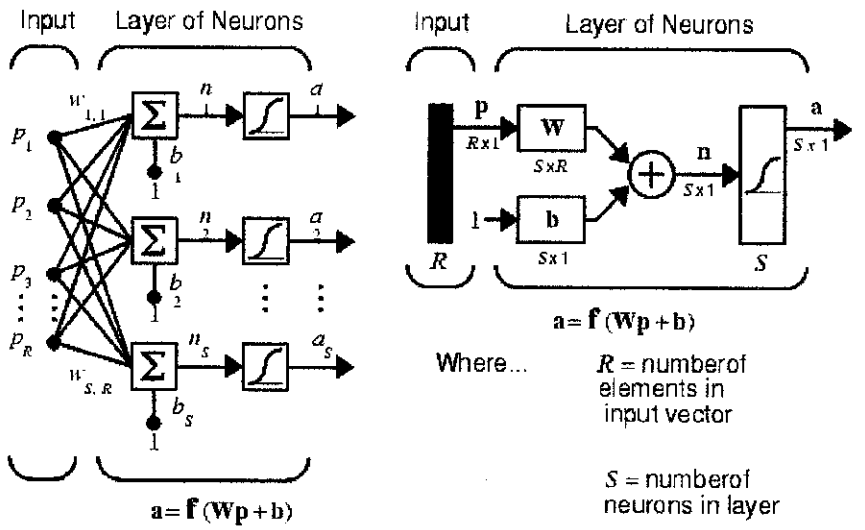



Figure 10 A single-layer networks (left) of S logsig neurons having R inputs and a layer diagram (right)

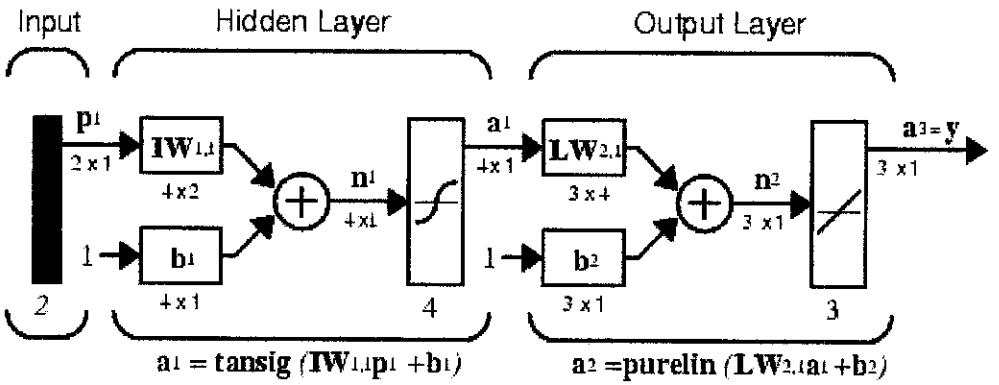



Figure 11 Basic backpropagation network

## 2.6 Detection of breast cancer

Mammography has a low specificity. The likelihood that a lesion found by mammography and sent to biopsy will be malignant is only 20 to 35% [16]. Numerous researches have been done to improve the detection of breast cancer by various methods and techniques. Strickland and Hahn used multiscale matched filters with wavelet transforms for enhancing and detecting calcification, Nishikawa et al use a difference technique to enhance microcalcification, and Moti Melloul and Leo Joskowicz use entropy tresholding in segmentation of microcalcification in X-ray mammograms [8].

Moti Melloul and Leo Joskowicz describe an algorithm that detects microcalcifications in two steps which removes background tissue with a multiscale morphological operation, and applies entropy tresholding based on a 3 –dimensional co-occurrence matrix. They use top-hat morphology to eliminate background tissue. They obtained mean detection rates of 93.75% of true positives, 6.25% of false positives, and 2.0% of false negatives. [8]

Armando Bazzani et al investigate the performance of a Computer Aided Diagnosis (CAD) system for detection of clustered microcalcifications in mammograms. They combined a multiresolution analysis based on wavelet transform with a difference-image method and gaussianity statistical test and they perform a logical OR operation on the detected microcalcification before clustering. [9]

Classification of clustered microcalcifications using fractal analysis and probabilistic neural networks by Wan Mimi and Diyana W Zaki and Rosli Besar proved that the probabilistic neural network are efficient for classification of clustered microcalcifications and manage to give reliable results for every mammogram tested.

They used standard deviations, first and second order entropy of the fractal thresholded images as input vectors. Clustered microcalcification is separated from breast background by their texture properties by fractal analysis. According to them, to extract feature, input layer must be verify first then select the design of neural network. To find the most suitable features, the thresholded fractal images

are analyzed and features are obtained mathematically. Then features are evaluated on a region of interest, and they identify the best variable values which are standard deviation, first order entropy, and second order entropy. [10]

Khairul Nisak Md Hasan [15] in her work titled Detection of Microcalcification using Mammograms enhanced the mammograms image by applying image processing technique using MATLAB and Borland C++. Top-hat algorithm method is developed using MATLAB. The method consists of digitization of mammograms, image enhancement, image segmentation, and feature extraction.

# CHAPTER 3
# METHODOLOGY

## 3.1 Procedure identification

Methodology used in the progress of this project includes information gathering through research on internet, books and journal, and also by interviewing experienced people in the medical field. As this is a two semester project, the tasks have been divided into two parts. This project is to design, develop, and test whether it run as expected. And the final process is to evaluate the output in order to analyze all tasks carried out from each phase. Project timeline can be view in APPENDIX II.

## 3.2 Project design

Figure 12 illustrate the steps involve in the project:

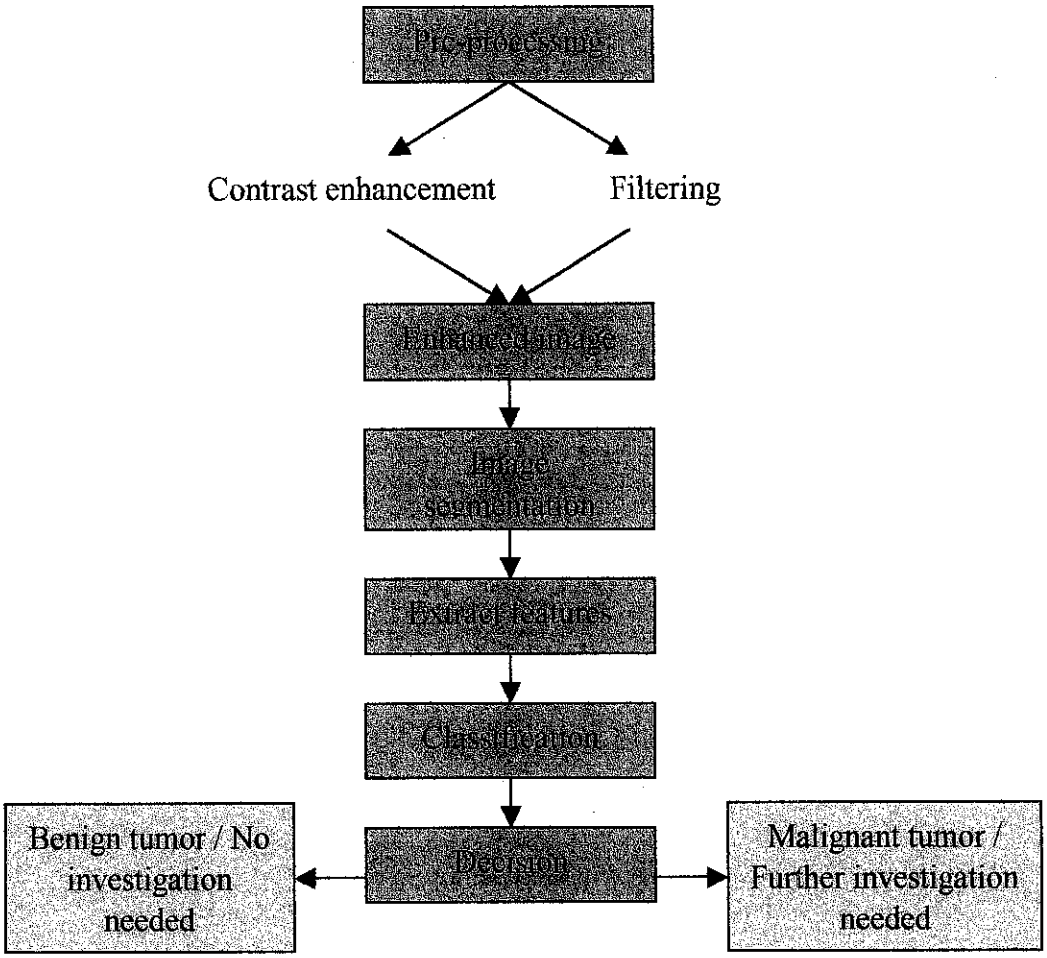



Figure 12 Steps involve in project development

## 3.3 Project development

### *3.3.1 Preprocessing*

Basically, preprocessing technique consist of gray scale manipulation, isolation of regions, noise filtering, contrast enhancement, image thresholding, and edge detection. All the sample of mammograms obtained from the hospital need to go through the process mentioned above before applying the higher level process.

In image acquisition, the mammograms sample is digitized using high resolution scanner, and stored in computer as '.tif' because the image should be in TIFF format in order to process it using MATLAB.

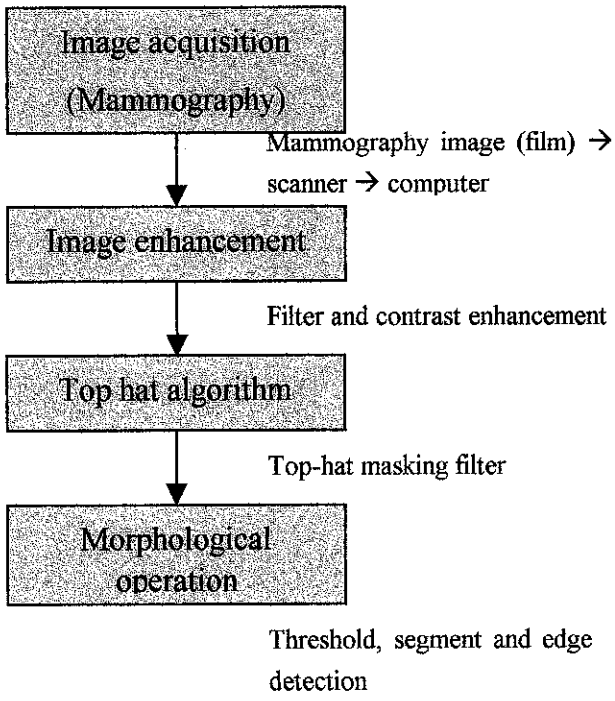



Figure 13 Image processing process

As indicated in Figure 13, to enhance the image, the first step is to remove noises. Gaussian low-pass filter operates as a smoothing mechanism to reduce noise. This filtering process results in an image with reduced "sharp" transitions in grey levels. Median filter is an order-statistics filter [12], [14]. It replaces the value of a pixel by the median of the grey levels in the neighborhood of the pixel, and it is better in reducing random noise without reducing the sharpness of the image.

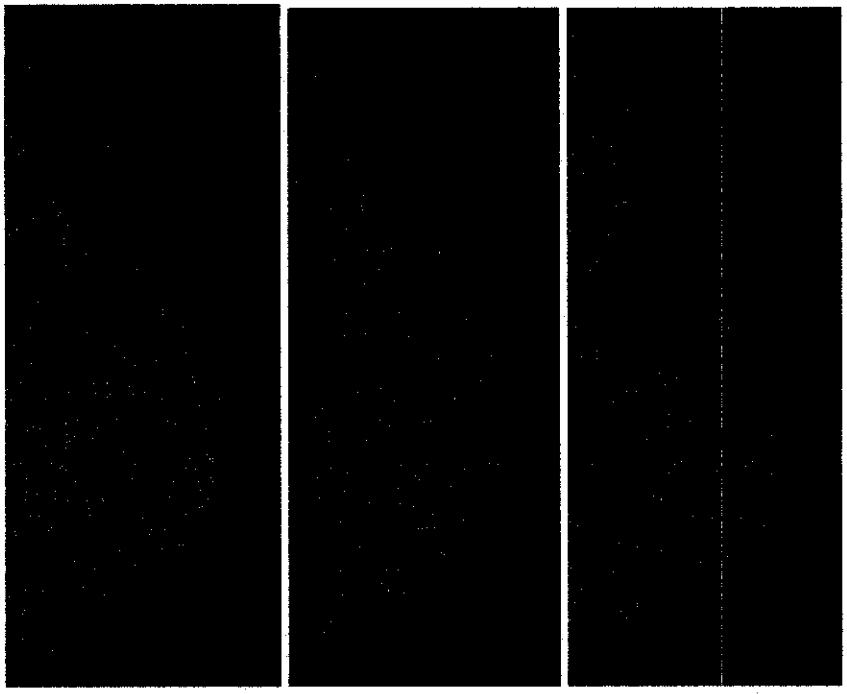


Figure 14 Effect of Gaussian filter and median filter (from left to right: original, Gaussian filtered image, and median filtered image)

However the sharp transitions in grey levels also consist of edges which are the advantageous features in an image, but averaging filters have the undesirable side effect that they can blur edges [14]. So to overcome the problem, unsharp masking filter is performed. This filter has the effect of making edges and fine detail in the image crisper and this approach is called high-boost filtering. Then intensity adjustment is performed to enhance the contrast of the image. The effect of each filter can be seen in Figure 14 and 15.

Top-hat masking filter with a disk-shape structuring element have been apply to the image to remove the uneven background illumination, and as the output of the operation is dark, we apply the *stretchlim* which calculates the histogram of the image and determines the adjustment limits automatically. Figure 16 showing the effect of *stretchlim* to top-hat filtered image.

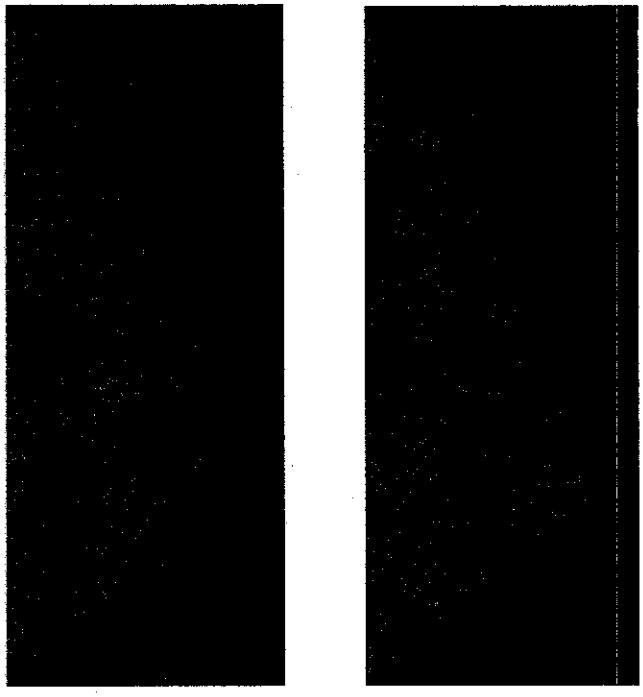


Figure 15 Image after applying unsharp masking filter and intensity adjustment

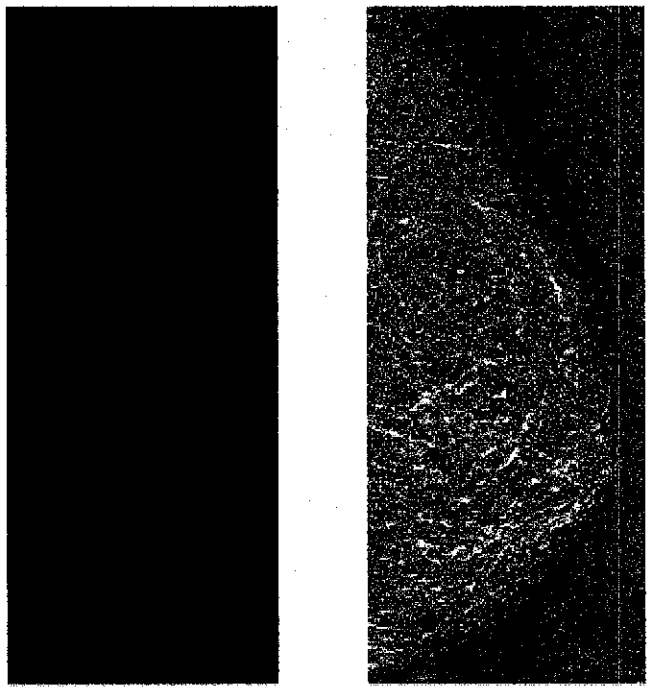


Figure 16 Image after applying top-hat filter and *stretchlim* function

After filtering and enhancing the contrast, image is threshold and erosion and dilation is performed. Erosion process is done to eliminate the boundary points from an object, leaving the object smaller in area by one pixel all around its parameter [12] (remove the unwanted small spot, artifacts).

Dilation is the process of incorporating into the object all the background points that touch it, leaving it larger in area by that amount [12] (in order to restore back the shape and size of the remainder). Figure 17 illustrate the effect of the erosion and dilation operation and the output of the preprocessing stage can be seen in Figure 18.

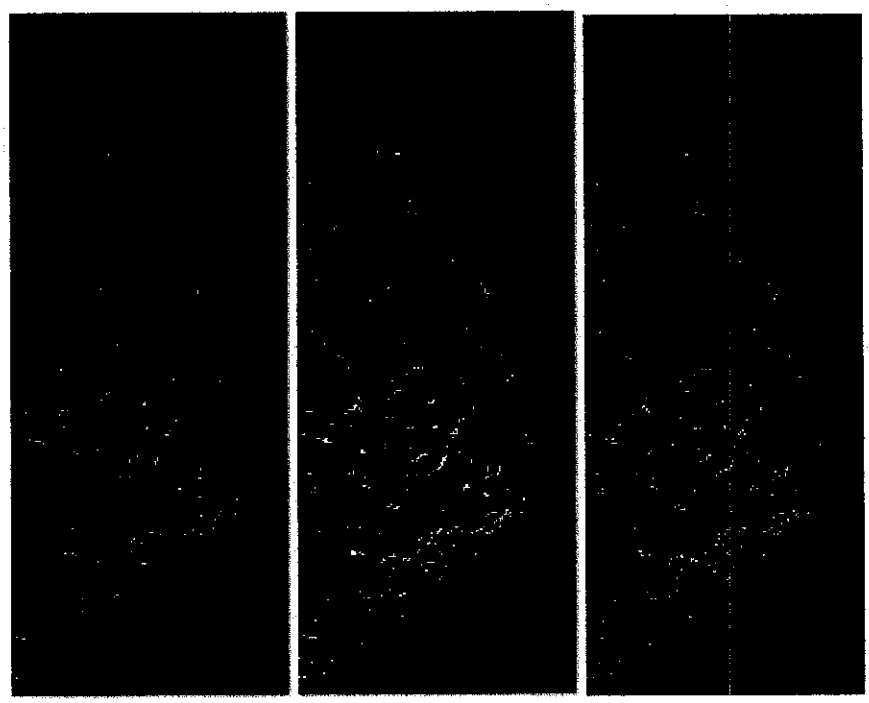


Figure 17 From left: segmented image (erosion), dilate gradient mask, and perimeter determination

Figure 18 Detected calcification

### 3.3.2 *Feature extraction*

After completing the processing stage, the next step is features extraction. The process of it is as Figure 19. It is performed on the binary image. Process of extracting the features have been divided into 5 steps which each steps presenting a feature. As the number of data is not enough for 5 process variables, the selection of appropriate features is needed. From the rough estimation, the features have been narrowed down into 4 features. The selected features are shape (circularity), eccentricity and orientation, and area.

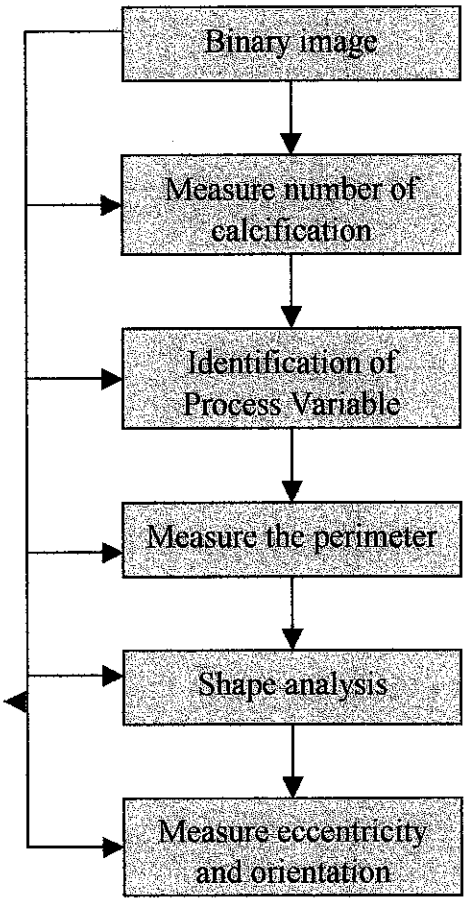



Figure 19 Process of feature extraction

### 3.3.3 *Neural network development*

Figure 20 shows the steps involve in the second part of the project. The steps consist of identification of process variable, data processing, neural

network construction and training, neural network validation and testing, interference of error, and the implementation of neural network.

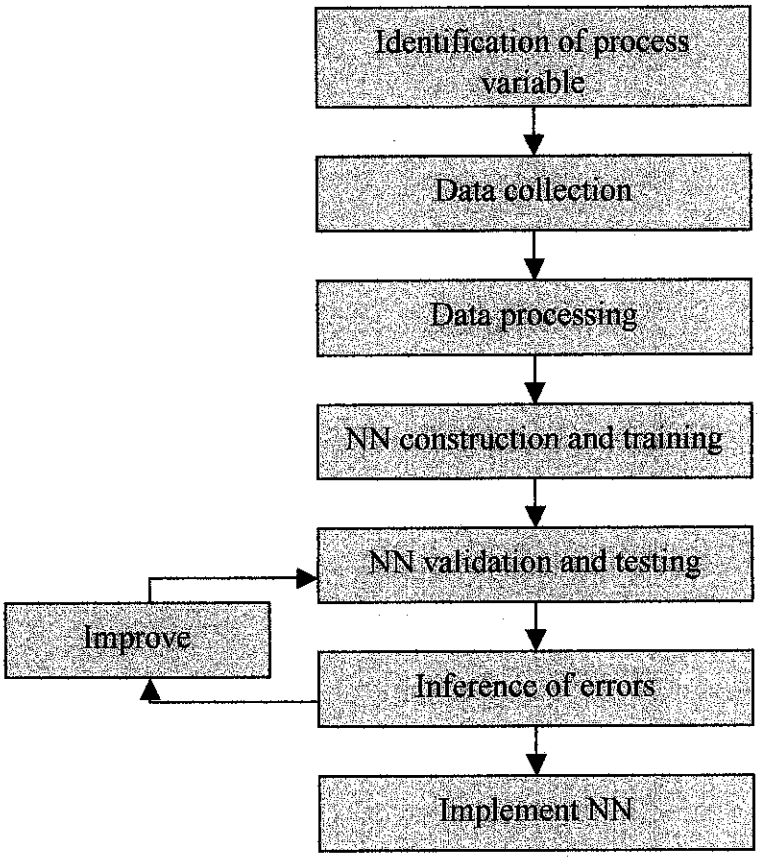



Figure 20 Methodology in developing the Neural Network (NN)

*Identification of process variable*

The neural network needs to be trained with sufficient amount of inputs and targets for it to be able predict. Process variable related to this study are the detection of microcalcification and the classification of it.

Sixty seven sets of data have been extracted, and 45 sets of the data are used to train the network. Thus the input variables can be set to 4, with estimation of 2 hidden layers and 1 output. Three sets of data that is training, validation and testing are generated using Microsoft® Excel's 'Random Number Generation'. Figure 21 illustrates the inlet and outlet process variables.

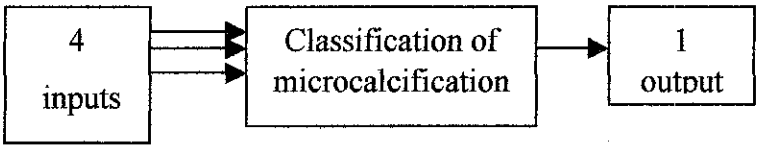


Figure 21 Pattern recognition process variables

*Preliminary Processing of Data*

Processing of data is done on set of inputs with its corresponding output. These work needed to be done before attempting to train the network. The sets of 45 inputs and output data needed to be divided to three sets that are training, validation and testing. Each set consist of right and left view of breast thus made up 90 input data. These sets are needed for different stages of work in neural network. The ratio between each set is according to the journal by Radhakrishnan and Mohamed (2000) that is 43% for the training, 43% for the validation and 14% for the testing. From that ratio, the training set has 39 sets of data, validation 38 sets of data and testing 13 sets of data.

Segmentation is conducted randomly using Microsoft® Excel's 'Sampling'. The software required user to specify the set of data for sampling and amount of sample size required. Sampling is done in all data sets. The random numbers generated is used in segmentation. The specified size for sampling must be larger than the desired size because the software replaces the number after selection. If sampling has repetition, the following sampled number is selected (the sets should not have repeated values).

After the segmentation, an ANOVA test is required to verify the original set and the three segmented sets are from the same population by comparing their means and standard deviations. The Microsoft® Excel's 'ANOVA: Single Factor' is used for this purpose. Test is conducted on the random number of the all data, training data, testing data and validation data. The means and standard deviation are compared.

*NN construction and Training*

There are three sets of inputs and output that arranged in a matrix form that are training set, validation set, and testing set. For the training sets the matrix arrangement is 4X39 for the inputs and 1X39 for the output. For the validation sets the matrix arrangement is 4X38 for the inputs and 1X38, and for the testing sets the matrix arrangement is 4X13 for the inputs and 1X13 for the output. MATLAB Neural Network Toolbox's 'Network/ Data Manager' (Figure 22) is used for constructing and training the network.

Figure 22 illustrate the network manager that is used to manage the neural network with the input and output. The data sets need to be load into the workspace before importing the data in the network/data manager. In order to create network, 'New Network...' button is used, and the window for creating a network is shown in Figure 23.

The proposed network used is 'Feed-forward Backpropagation'. The input range should be specified and it can be obtain from training inputs. To find the suitable neural network configuration, it can be determined by changing the training function, adaptation learning function, performance function number of layers, number of neurons and transfer function (trial and error). The desired output in the study must be positive in value, therefore the last layer utilized the transfer function of *logsig*. The performance curve of the network needs to be analyzed to identify the suitable configuration that produced the minimal error.

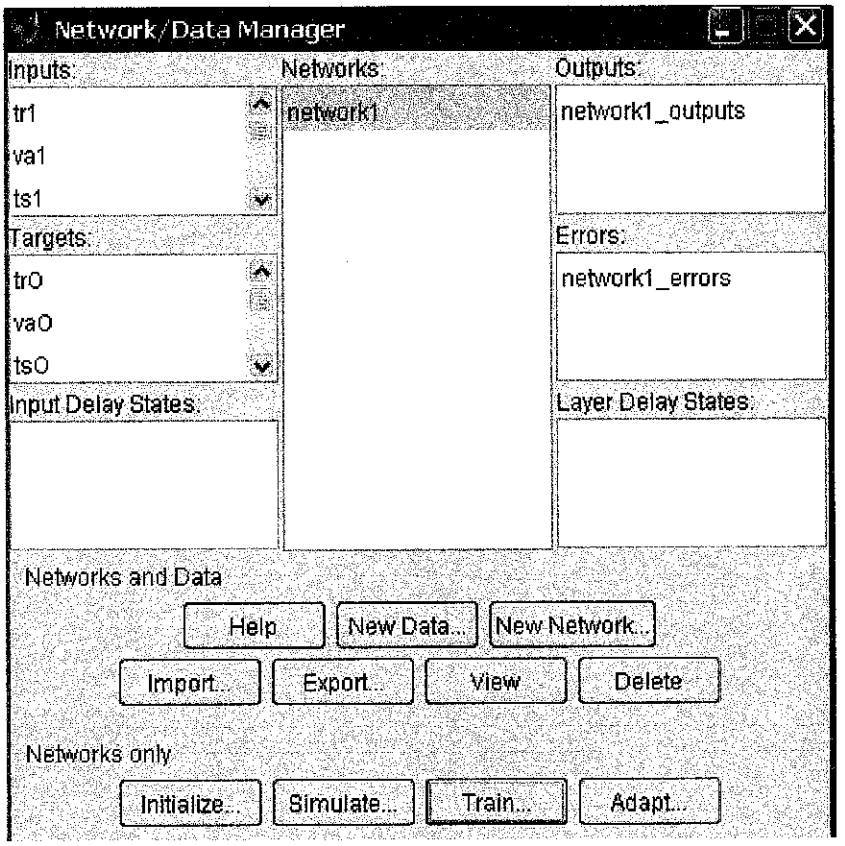



Figure 22 Neural Network/Data Manager

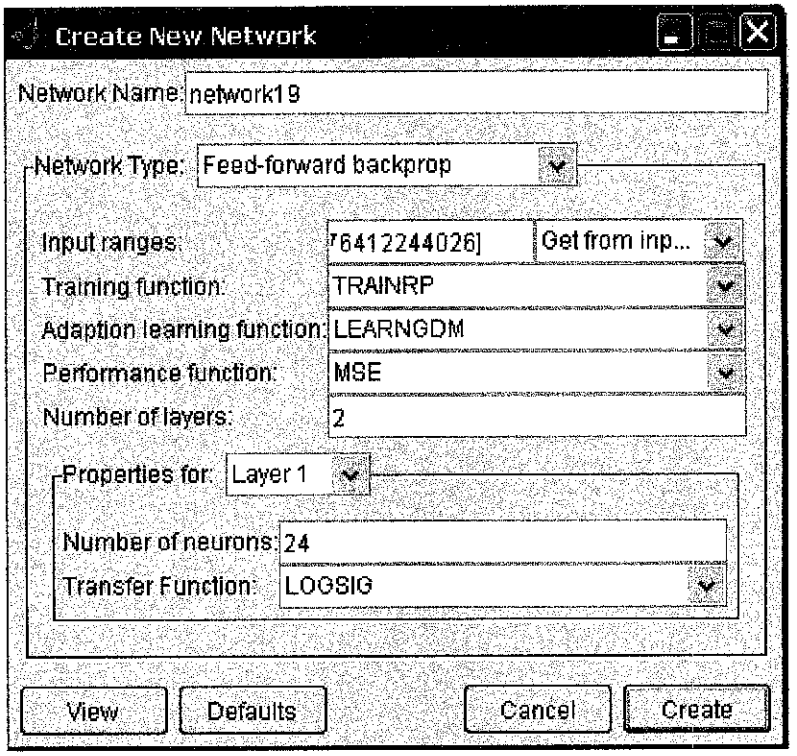
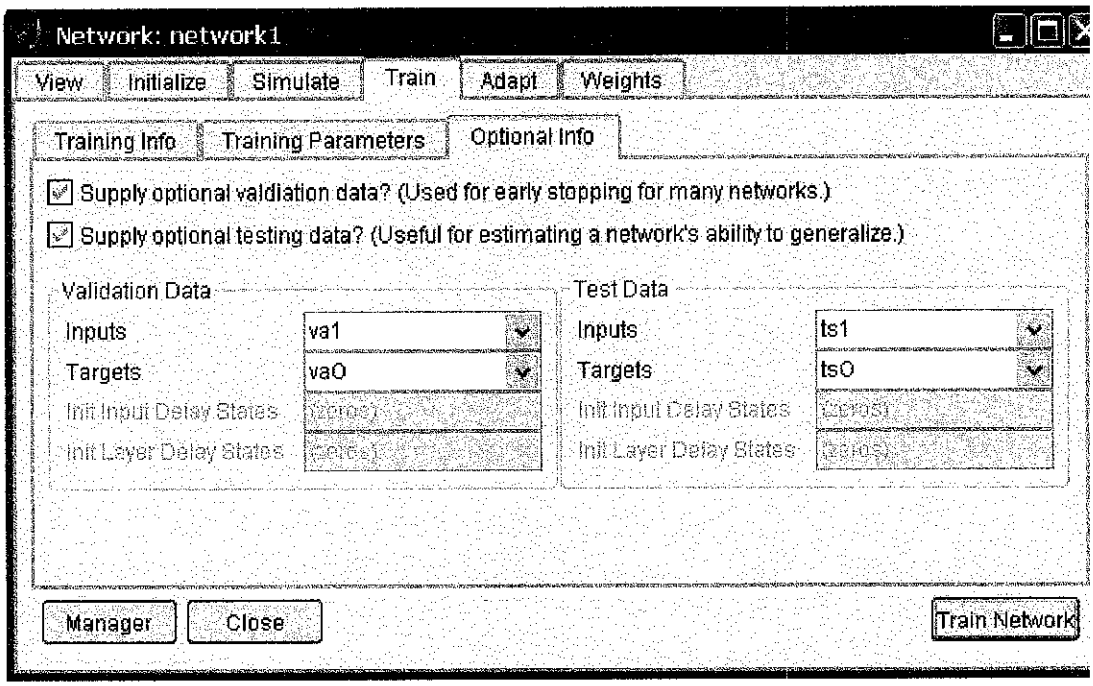



Figure 23 Create New Network

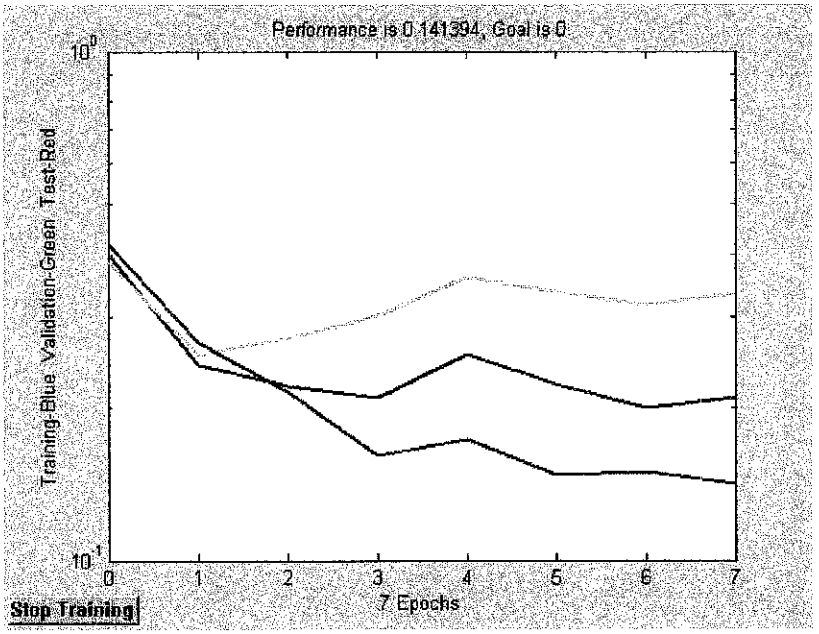
*NN Validation and Testing*

In determining a suitable network, the validation and testing set must be used with the training set so that a reasonable configuration network can be identified. It is used by supplying these sets of data before training shown in figure, so that an approximation performance curve for all sets can be generated as shown in Figure 24.

Thus additional information provided as Figure 24(a) will generate the curve as in Figure 24(b). This is useful for classifying a suitable network. The testing set is not simulated to obtain the output. The reason it is useful is that the curves of the validation and testing set must be below the training set as one of the criteria for the optimum configuration. If the curves of the validation and training are higher than the training set, the error generated is much higher than the training set. Hence it is required to determine a configuration that produces validation and testing curve below the training curve. Otherwise the network is not able to generated is robust and accurate prediction.




(a)




(b)

Figure 24 (a) and (b): Consideration for construction of network

After a suitable configuration is identified, the validation set is used for validating the network in its performance by simulating it using the trained network created. If the results are satisfactory, testing can be conducted using the testing set. If not the network must be retrained with different configuration until it is successful.

*Testing of Error*

Error testing is conducted only on validation and testing set. Error is calculated on the Root Means Square Error (RMSE). RMSE determined the error between the predicted and actual plant values, square them, sum them, divide by the number of the data point and determined the square root of them.

*RMSE = sqrt (sum ((predicted value-actual value) ^2)/number of data)* (5)

Potential improvement is done after suitable neural network is constructed. The purpose this is to further minimize the error in prediction value. The error for the best modeling is must be less then 5%.

# CHAPTER 4
# RESULT AND DISCUSSION

## 4.1 Result

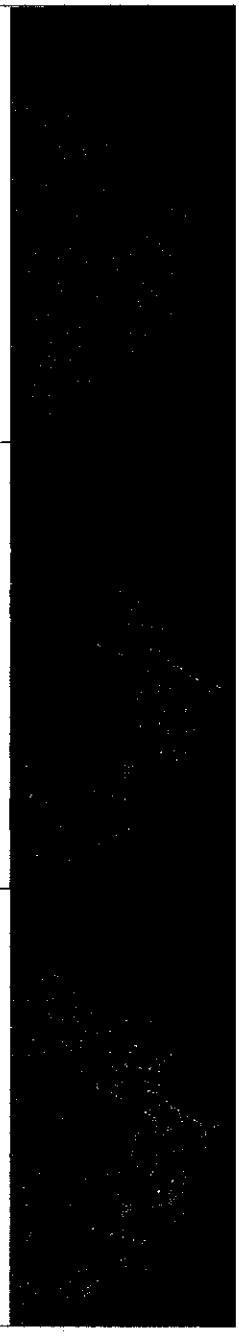
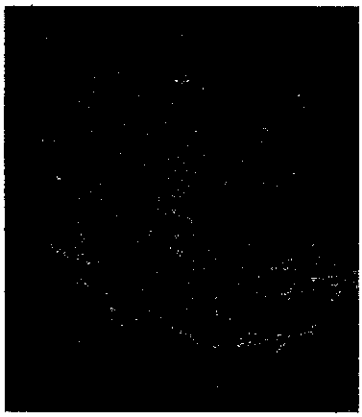
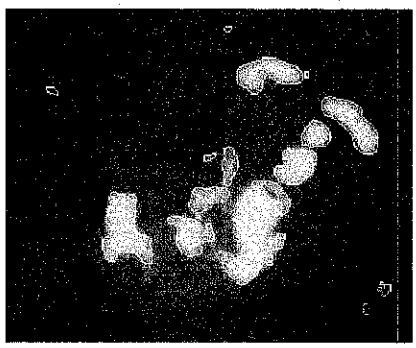
The preprocessing process, which is filtering, and contrast enhancement, and image segmentation have been done to the 88 cases by the method discussed in the previous chapter. Which 23 cases were taken from the previous project [15], 20 from Hospital Ipoh, and another 45 cases was downloaded from the "Interactive Mammography Analysis Web Tutorial" [23]. The cases that was obtained from the internet and hospital have their details, but not to the remaining 23 cases. The mammogram samples were processed using the program in APPENDIX C. Sample of three results that have been processed by using MATLAB with opening morphological operation is as in Table 1.

Features have been extracted from 67 cases and the 45 cases downloaded from the internet is used for neural network construction and training. The result of feature extraction is as in APPENDIX D. The extracted features are number of calcification, area, perimeter, eccentricity and orientation, solidity, and convex area. Circularity formula (3) is computed by using average area and average perimeter. Area and perimeter units are in pixels. Figure 25 illustrates the result of feature extraction of 45 cases.

Result of preliminary data processing is attached in APPENDIX D. The output of segmentation is three sets of data that are 39 data for training, 38 data for validation and 13 data for testing. The segmentation was done using random number of inputs data. ANOVA test was performed to verify the original and the segmented sets are from the same population. The test was performed on the random numbers of the input and output variable. From the result of ANOVA test, it can be seen that the average value of segmented sets is near to the average

value of the original set. Thus it concludes that the segmented sets are from the original set.

Table 1 Sample result of three cases (detection of calcifications)

| **Description/ samples** | **Image 1** | **Image 2** | **Image 3** |
|---|---|---|---|
| Original |  |  |  |
| Perimeter determination |  |  | |
| Detected calcification |  |  | |
| No. of calcification | 36 | 24 | 61 |
| Total area | 1843.3333 | 2176.3333 | 2375.6667 |
| Mean area | 51.2037 | 90.6806 | 38.9454 |
| Perimeter | 955.6667 | 1443.3333 | 1209 |

Numerous trail and errors have been performed to obtain the optimum configuration and the most suitable configurations that can be obtained for the neural network are as in Table 2, and the curves in Figure 26 indicate the performance of the neural network in predicting all the three sets. Desired error is 0 and the performance is 0.141394. Amount of iteration (epoch), for predicting the value of performance is 7. The performance is with regards to the termination due to validation and testing set. The actual performance based on training set is 0.0992984 with termination at 100 epochs. This is the result that produces the minimal error.

Table 3 shows sample of result of testing. The input of testing set is simulated and the predicted result by neural network and the error of the prediction is tabulated. The graph in Figure 27 illustrates the performance of the prediction value. The actual output or result of the test is 0 or 1 which 0 indicates no further investigation needed, and 1 indicates further investigation needed.

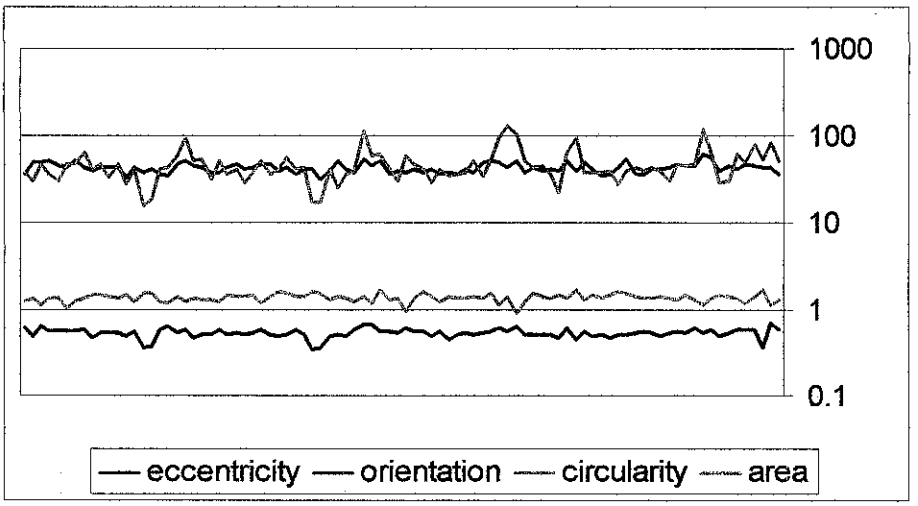



Figure 25 Results of feature extraction

Table 2 Network configuration

| Parameters | Variable |
|---|---|
| Network | Feed-forward backprop |
| Training function | TRAINRP |
| Adaptation learning function | LEARNGDM |
| Performance function | MSE |
| Epochs | 100 |
| Number of layer | 2 |
| Layer 1: Number of neuron | 24 |
| Transfer function | LOGSIG |
| Layer 2: Number of neuron | 1 |
| Transfer function | LOGSIG |

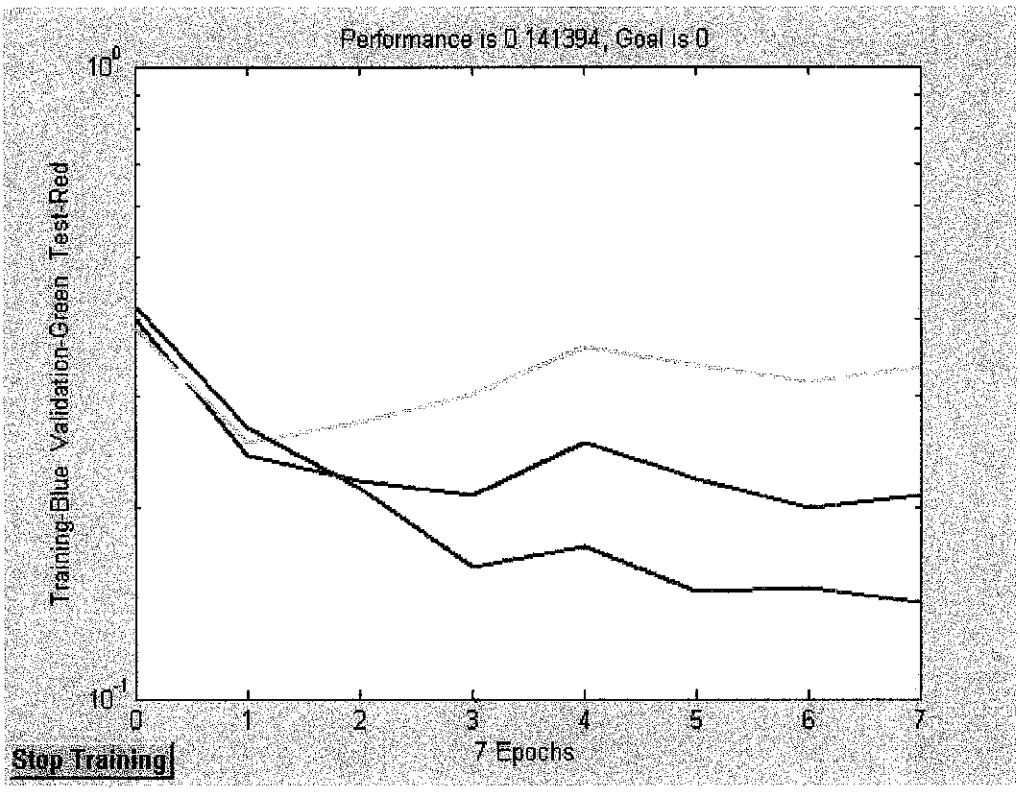



Figure 26 Performance curve of the network

Table 3 Sample of result predicted by Neural Network

| Actual value | Predicted value | Error |
|---|---|---|
| 0 | 0.1582 | 0.1582 |
| 1 | 0.49987 | 0.5001 |
| 1 | 0.98054 | 0.0195 |
| 1 | 0.70959 | 0.2904 |
| 1 | 0.66996 | 0.33 |
| 0 | 0.89372 | 0.8937 |
| 1 | 0.50595 | 0.494 |
| 1 | 0.15339 | 0.8466 |
| 0 | 0.19851 | 0.1985 |
| 1 | 0.9687 | 0.0313 |
| 1 | 0.64472 | 0.3553 |
| 0 | 0.089263 | 0.0893 |
| 0 | 0.15175 | 0.1517 |

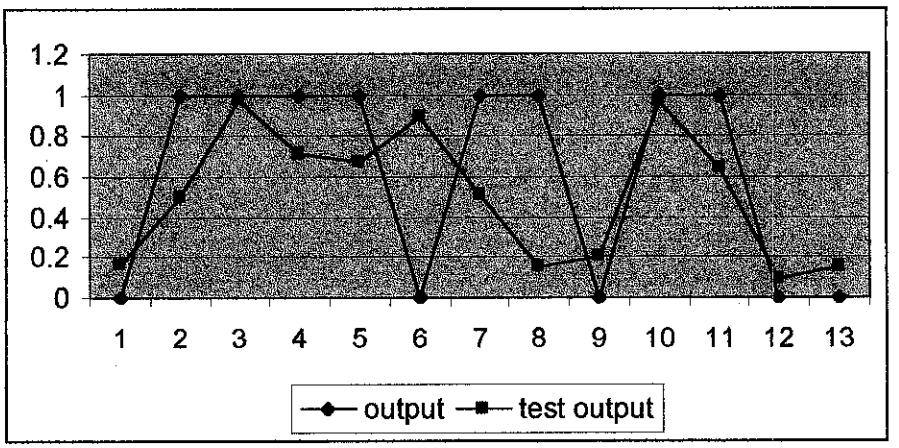



Figure 27 Graph of actual value vs. predicted value

## 4.2 Discussion

The steps taken in each stage of preprocessing and feature extraction have been discussed in the previous chapter. Since this project is a continuation from the prior project [15], some of the preprocessing steps have been used. The existing program coding has been modified to enhance the output.

There was few problems occur during developing the coding in MATLAB. One of the problems was on the output produced from the preprocessing stage. The output is not constant. By observation, the output is proportional to the cropped area as most of the operation involving the averaging of grey-level values. The effect of cropping can be seen in Figure 28 and 29.

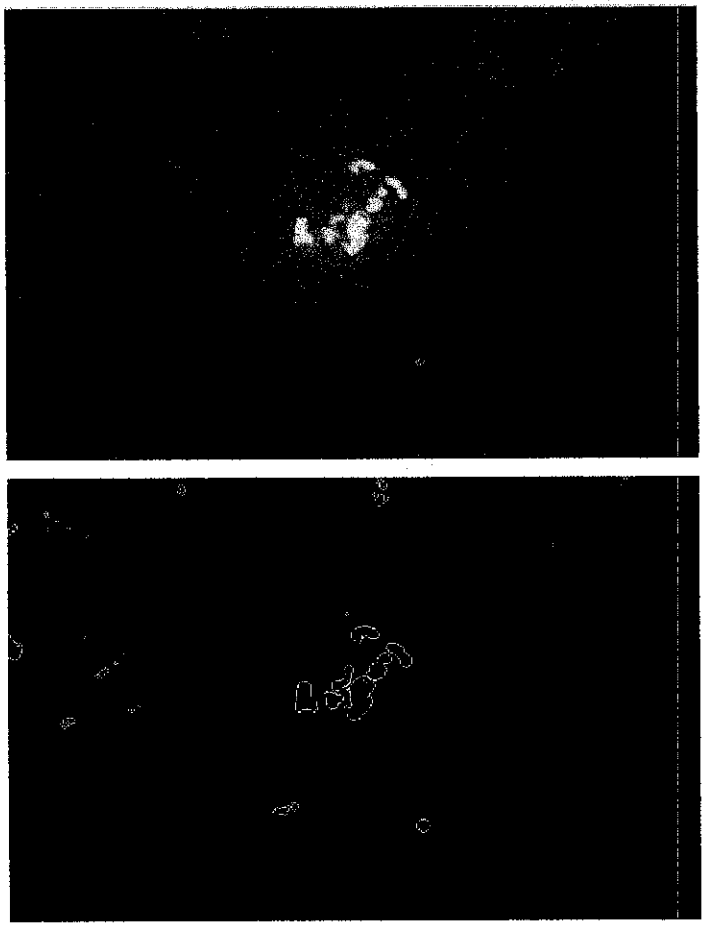


Figure 28 Effect large cropped area (above: cropped image, below: output)

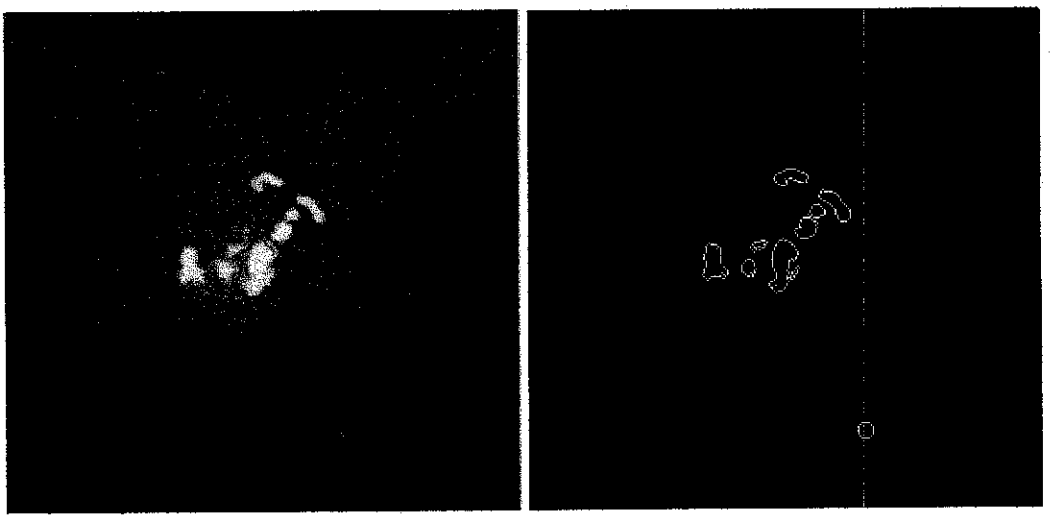


Figure 29 Effect on small cropped area

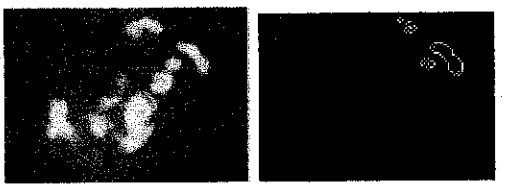


Figure 30 Effect on smaller cropped area

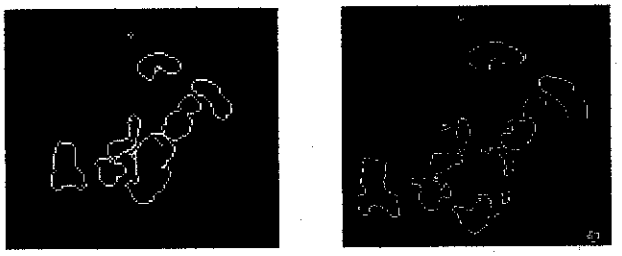


Figure 31 Cropped area of interest of Mammo20-1 (left: resized image, right: image without resizing)

As in Figure 28, the image of breast is wholly cropped, and the image in Figure 29 is only focused on the interested area, and the difference can be seen clearly when the image is cropped to focus on the suspected area (as in Figure 30). In order to have a convenience experimental value each of the images can be processed three times and the results will be taken from the average value of it.

The processes of image enhancement have been completed and the resizing effect has been tested. Figure 31 shows the resizing effects on the detected area. Both of the images have same masking size, thus affecting the detected area. To put it

briefly, resize to smaller image size will not reduce any detail or data in the image, unlike image enlargement.

To obtain the results as in APPENDIX D (Result of feature extraction), each of the images have been processed through image processing technique and feature extraction method. The extraction stage has been verified by testing a nearly circular object, and an irregular object that is taken from one part of the mammograms (see APPENDIX E). The MATLAB coding can be view in APPENDIX C. The circularity is calculated by using equation (3). Figure 32 shows the sample of mammogram used for the verification of circularity, and circle objects and irregular object taken from the sample are as in Figure 33 and 34.

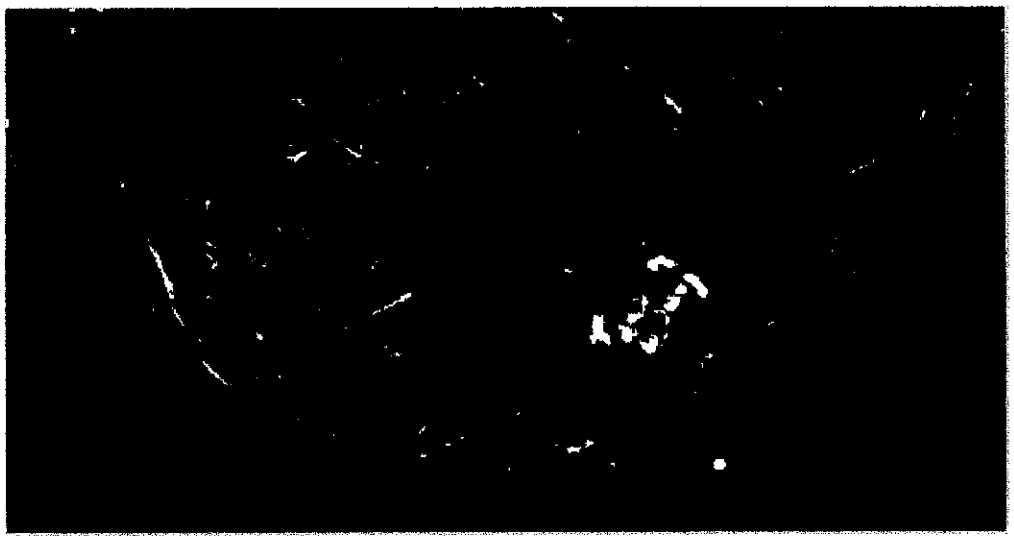


Figure 32 Detected calcification of mammo20-1

Figure 33 Circle object cropped from mammo20-1

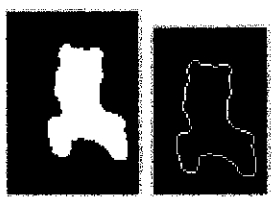


Figure 34 Irregular object cropped from mammo20-1

The average area of Figure 33 is 442 pixels. The results of feature extraction for the object are:

- meanArea = 442
- meanEccentricity = 0.1412
- meanOrientation = 17.6378
- meanSolidity = 0.9736
- meanPerimeter = 68
- meanCircularity = 1.2012

And the average are of figure 34 is 1387 pixels. And the results are:

- meanArea = 1387
- meanEccentricity = 0.7906
- meanOrientation = -75.7992
- meanSolidity = 0.7642
- meanPerimeter = 173
- meanCircularity = 0.5824

The result of computed feature extraction is compared with the result of manually extracted which is the data base of the mammogram. Features that have been extracted and are used as the input data are circularity, eccentricity, orientation, and area. Forty five sets of data are taken from the internet and twenty two sets of data are taken from Hospital Ipoh. Altogether are 67 sets of data that consists of 4 inputs and 1 output. Details about the output were attached in APPENDIX D (Result of preliminary data processing).

The sets of 45 data have been divided into three sets that are training, validation and testing. Each set consist of right and left view of breast thus made up 90 input data. The ratio between each set is 43:43:14; training set has 39 of data, validation 38 of data and testing 13 sets of data. ANOVA test have been done to the segmented data. Test is conducted on the random number of the all data, training

data, testing data and validation data. The means and standard deviation are compared.

The neural network was developed with configuration of Table 2. The network that was generated by using the configuration is shown in Figure 35. There are two layers as indicated by the block with the numeric at the bottom. The first layer has 24 neurons. The inputs are connected to the nodes in the input layer. The output layer has 1 neurons and the output is taken from this layer. The performance curve during training configuration is shown in Figure 36. The performance obtained is 0.0992984, which is the closest to the desired error of 0 with the iteration of 100 times. The performance is quite poor as the desired performance is approximately 0.001 or less. This is due to small number of training set. Thus the performance can be enhanced by increase the number of data set.

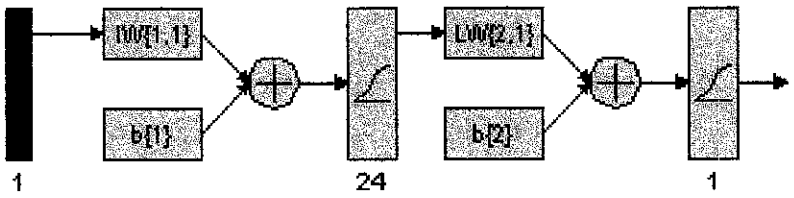



Figure 35 Network generation

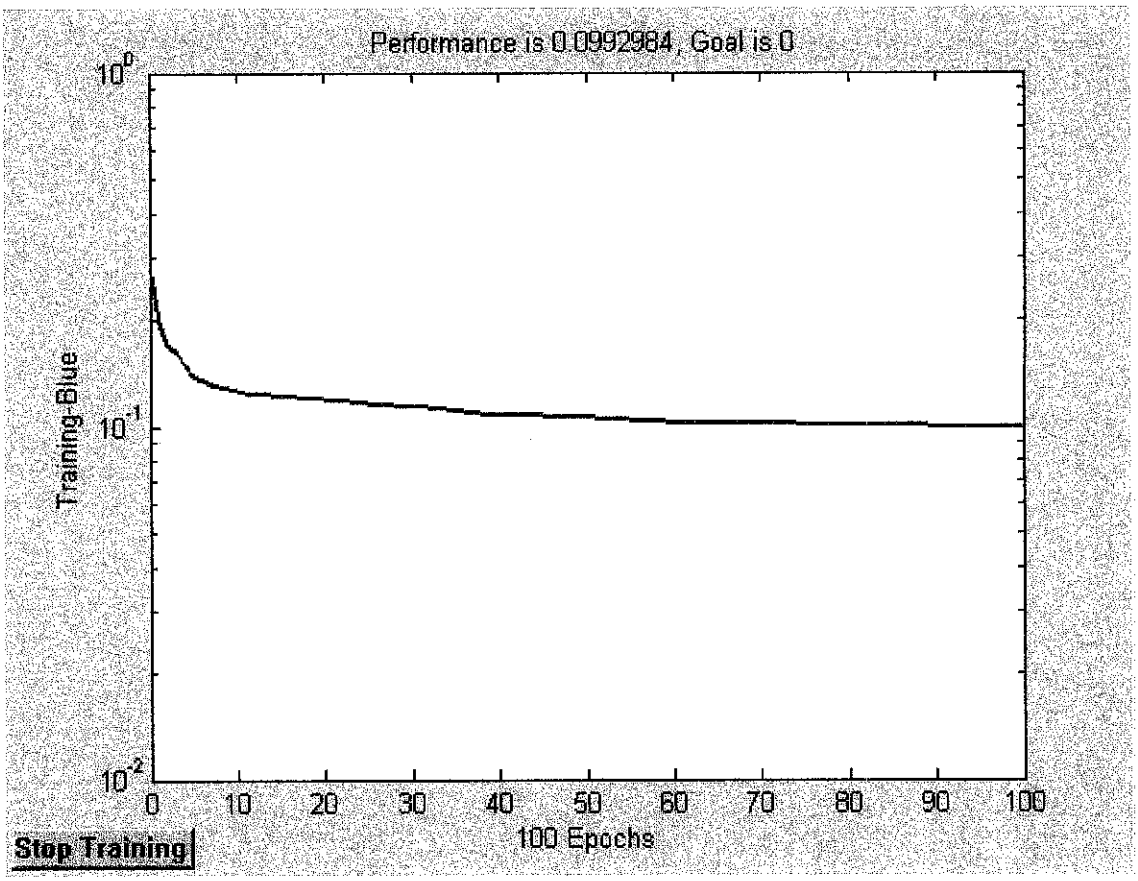



Figure 36 Performance curve during training

Testing was conducted to ensure that the configurations are suitable for prediction. The graph of the performance is shown in Figure 27. As can be seen, the predicted output is lack in accuracy. This is due to the poor performance of training.

The neural network simply predicts the performance using the input of the testing set. This able the network to measure the error generated. The error is high (as can be seen in Figure 37) due to the lower number of iteration, which are 7. This is because the iteration terminates at $7^{th}$ iteration compared to the training set (100 iteration). The effect is due to failing of other sets input to converge which cause early termination.

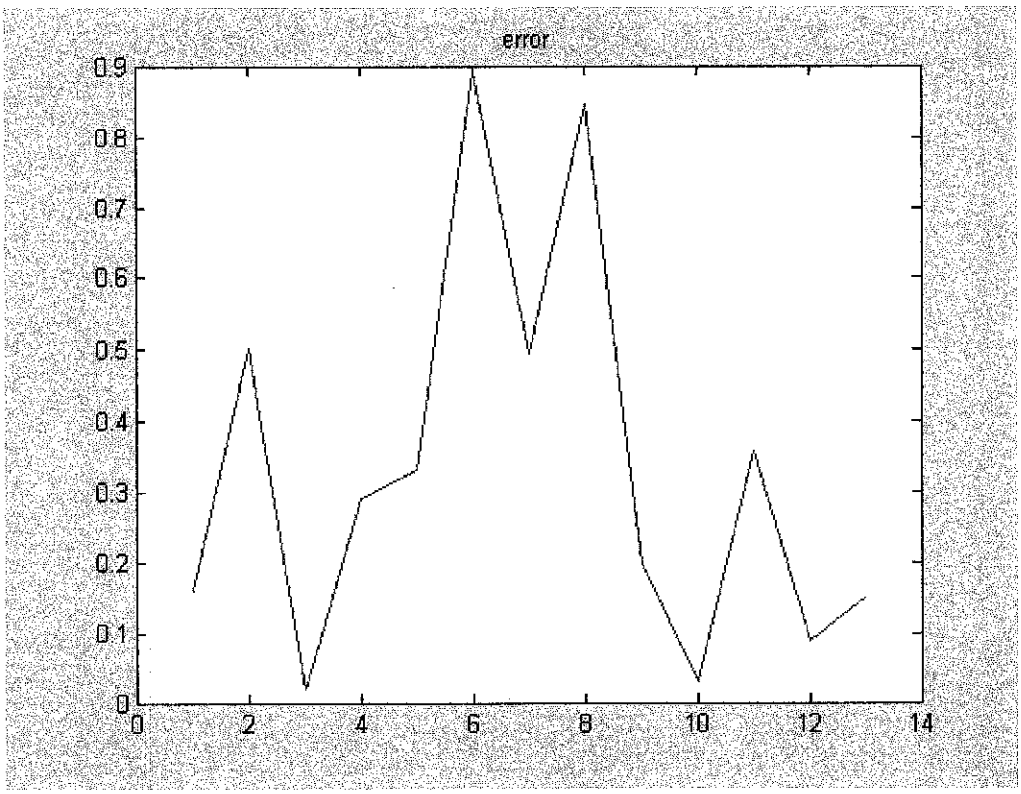



Figure 37 Generated error

From this project, it had been determined that there is inconsistency between number of neuron and performance. Error is smaller when neuron number is increased. However, higher number of neuron may decrease the number of iteration to achieve the performance thus increasing the error. Generally when the curve is of decreasing nature due to the higher number of iteration, the convergence is higher, error is smaller and the offset to the desired error is smaller.

From the result obtained by prediction of test set, the network is capable of classifying the result with RMSE of 0.3816, which is 38.16%. And a good modeling of neural network required the RMSE to be less than 5%. The problem in achieving the good modeling must be because of the performance of the input data and not enough data set in training the network.

# CHAPTER 5
# CONCLUSION AND RECOMMENDATION

## 5.1 Conclusion

According to all the information that has been stated earlier, mammogram images have been enhanced through image processing technique using MATLAB. Opening morphological operation has been used to detect the calcification. From that, features have been extracted from binary image and the most significant features are used for the classification stage.

As the result, a pattern recognition system has been provided by using neural network. But due to the large error and poor performance, this system has not met the objective. The problem and recommendation to this error will be discussed further in the next section.

The accomplishment of this project could help medical field in detecting breast cancer.

## 5.2 Recommendation

As the project has not met the target, some possibilities of error might involve throughout project. These potential errors are due to human and system errors. However, the occurrence of these errors can be reduced by taking the pre-cautions steps, and the project can be enhanced by some of these recommendations below:

i. Further study on the detection of breast cancer should be done for the minimization of false positive and false negative. This can be done through research on various methods in detecting the abnormalities. For example, region growing algorithm (automated region of interest), image

tresholding, and various segmentation algorithm such as watershed algorithm.

ii. Since the system error is because of lack in source, increasing the number of case study (mammogram image) is necessary. But need to put into constraint that each case study must have their details record of the case for further use in neural network construction and development.

**iii.** In order to enhance the performance of input data in neural network, significant features that emphasize the difference between benign and malignant case are needed. As an example, D Betal et al [24] suggested to use numerical analysis of segmented microcalcification to distinguish between benign and malignant clusters by using shape analysis, cluster features analysis, and receiver operating characteristic (ROC) analysis. Some of the essential and valuable features for classification are texture feature, shape feature, and scalar area features [20].

**iv.** Finally, the project can be enhanced in many ways to provide a good pattern recognition system through a various stages and methods. The proper procedure must be consistent for all analysis performed. The strategy of training the network and research to improve the performance and error of neural network is needed.

# REFERENCES


1. http://www.nlm.nih.gov/medlineplus/ency/
2. http://www.mma.org.my/current_topic/women.htm
3. http://www.breastcancer.org
4. http://www.nci.nih.gov/cancertopics/wyntk/breast
5. http://www.cancer.org
6. http://www.personalmd.com/news/digital_mammography_041000.shtml
7. Margarita Sordo; Introduction to Neural Networks in Healthcare; for OpenClinical; October, 2002.
8. Moti Melloul and Leo Joskowicz; *Segmentation of microcalcification in X-ray mammograms using entropy tresholding*; The Hebrew University of Jerusalem, Israel.
9. Armando Bazzani et al; *Automatic detection of clustered microcalcifications in digital mammograms using an SMV classifier*; University of Bologna, Italy.
10. Wan Mimi Diyana W Zaki and Rosli Besar; *Classification of Clustered Microcalcification using Fractal Analysis and Probabilistic Neural Network*; Multimedia University, Malaysia.
11. Abhijit S. Pandya, Robert B. Macy; Pattern *Recognition with Neural Network in C++*; CRC Book Published in Cooperation with IEEE Press.
12. Kenneth R.Castleman; *Digital Image Processing*; Prentice Hall; Upper Saddle River, NJ; 1996.
13. Nick Efford; *Digital Image Processing-a practical introduction using Java*; Addison-Wsley.
14. Rafael C. Gonzalez and Richard E. Woods; Digital Image Processing; Prentice Hall; Upper Saddle River, New Jersey; second edition, 2002.

15. Khairul Nisak Md Hasan; *Detection of Microcalcification using Mammograms*; Final Year Project JUNE 2004, University Technology Petronas.
16. Fred S. Azar; Imaging Technique for Detecting Breast Cancer: Survey & Perspective; University of Pennsylvania; Philadelphia (PA), 2000.
17. Piotr Wasilewski; Image Processing.
18. Isaac N. Bankman et al, Segmentation Algorithms for Detecting Microcalcifications in Mammograms, IEEE Trans. on Info. Tech. in Biomedicine, Vol. 1, no. 2, JUNE 1997.
19. Hagan, Demuth, Beale, Neural Network Design, PWS Publishing Company, 1996.
20. A. Wróblewska et al, Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms, contributed paper, Opto-Electronics Review II (3), 227-235 (2003).
21. Milan Sonka, Vaclav Hlavac, and Roger Boyle, Image Processing, Analysis, and Machine Vision, second edition, PWS Publishing, 1999.
22. http://sprojects.mmi.mcgill.ca/mammography/masses.htm
23. V.R. Radhakrishnan and A.R Mohamed, 'Neural Network for the Identification and Control of Blast Furnace Hot Metal Quality', Journal of Process Control 10, 2000.
24. D Betal et al, Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology, The British Institute of Radiology, 1997.

# APPENDICES

# APPENDIX A

# STAGE OF BREAST CANCER

| Stage | Description |
|---|---|
| STAGE 0 | In Situ ("in place") disease in which the cancerous cells are in their original location within normal breast tissue. Known as either DCIS (ductoral carcinoma in situ) or LCIS (lobular carcinoma in situ) depending on the type of cells involved and the location, this is a pre-cancerous condition, and only a small percentage of DCIS tumors pregress to become invasive cancers. There is some controversy within the medical community on how to best treat DCIS. |
| STAGE I | Tumor less than 2 cm in diameter with no spread beyond the breast |
| STAGE IIA | Tumor 2 to 5 cm in size without spread to axillary (armpit) lymph nodes or tumor less than 2 cm in size with spread to axillary lymph nodes |
| STAGE IIB | Tumor greater than 5 cm in size without spread to axillary lymph nodes or tumor 2 to 5 cm in size with spread to axillary lymph nodes |
| STAGE IIIA | Tumor smaller than 5 cm in size with spread to axillary lymph nodes which are attached to each other or to other structures, or tumor larger than 5 cm in size with spread to axillary lymph nodes |
| STAGE IIIB | The tumor has penetrated outside the breast to the skin of the breast or of the chest wall or has spread to lymph nodes inside the chest wall along the sternum |
| STAGE IV | A tumor of any size with spread beyond the region of the breast and chest wall, such as to liver, bone, or lungs |

# APPENDIX B

# PROJECT TIMELINE

## Milestone for the First Semester of Final Year Project

| | Detail/ Week | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Selection of topic<br>- proposed topic<br>- topic assigned to student | | | | | | | | | | | | | | |
| 2 | Literature review<br>- breast cancer and mammograms<br>- methodology | | | | | | | | | | | | | | |
| 3 | Preliminary research work<br>- introduction<br>- objective<br>- list of reference<br>- project planning | | | | | | | | | | | | | | |
| 4 | Submission of preliminary report | | | | ● | | | | | | | | | | |
| 5 | Project work<br>- literature review<br>- learning process<br>- work on project (image enhancement) | | | | | | | | | | | | | | |
| 6 | Submission of progress report | | | | | | | | ● | | | | | | |
| 7 | Project work continue<br>- pre-processing process<br>- feature extraction | | | | | | | | | | | | | | |
| 8 | Submission of interim report final draft | | | | | | | | | | | | ● | | |
| 9 | Submission of interim report | | | | | | | | | | | | | ● | |
| 10 | Oral presentation | | | | | | | | | | | | | | ● |

Milestone ● Process done Expected process

## Milestone for the Second Semester of Final Year Project

| No. | Detail/ Week | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Project Work Continue<br>- Finalize preprocessing<br>- Finalize extracting feature<br>- selection of neural network method | ■ | ■ | ■ | | | | | | | | | | | |
| | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | | | | | | |
| 2 | Submission of Progress Report 1 | | | | ● | | | | | | | | | | |
| 3 | Project Work Continue<br>- design network<br>- train network | | | | ■ | ■ | ■ | ■ | | | | | | | |
| | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| 4 | Submission of Progress Report 2 | | | | | | | | ● | | | | | | |
| 5 | Project Work Continue<br>- test network<br>- finalize pattern recognition | | | | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| | | | | | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| 6 | Exhibition (Pre-EDX) | | | | | | | | | | | | ● | | |
| 7 | Submission of Draft Report | | | | | | | | | | | | | ● | |
| 8 | Submission of Final Report (soft cover) – study week | | | | | | | | | | | | | | |
| 9 | Submission of Technical Report – 13/5 | | | | | | | | | | | | | | |
| 10 | Oral Presentation – 6/6 - 8/6 | | | | | | | | | | | | | | |
| 11 | Submission of Project Dissertation – 24/6 | | | | | | | | | | | | | | |

Milestone ● Process done ■ Expected process ■

# APPENDIX C
# PROGRAM CODING

Preprocessing:

```
I=imread('mammo20-1.tif');          % Read image file
A=imcrop(I);                        % Select region of interest (crop)
B=imresize(A, 0.5, 'bicubic', 3);   % Resize the image to 75% of the
                                    % cropped image using
                                    % bicubic interpolation method by
                                    % 3-by-3 lowpass filter
figure,imshow(B),title('original')
figure,imhist(B),title('original')

h=fspecial('gaussian')              % Create Gaussian lowpass filter
A2=imfilter(B,h,'replicate');       % Perform Gaussian lowpass filter
                                    % to image B
figure,imshow(A2),title('Gaussian lowpass filter')
```

```
L=medfilt2(A2,[3 3]);          % Apply median filter using
                               % filtering size of 3-by-3
                               % neighborhood
figure,imshow(L),title('Median filter')
p=fspecial('unsharp')          % Returns 3-by-3 unsharp contrast
                               % enhancement filter
U=imfilter(L,p,'replicate');   % Apply an unsharp masking filter
figure,imshow(U),title('unsharp contrast enhancement filter')
Z=imadjust(U,([]),[0 1]);      % Maps the values in the intensity
                               % image U to new values in Z such
                               % that values between low_in and
                               % high_in map to values between low_out
                               % and high_out
figure,imshow(Z),title('intensity adjustment')
se=strel('disk',12);           % Create a morphological structuring
                               % element of disk-shape with radius 12
J=imtophat(Z,se);              % Perform top-hat filter to the image
figure,imshow(J),title('top-hat filtering')
K=imadjust(J,stretchlim(J),[]); % Increase the contrast of the image
```

```
figure,imshow(K),title('improved visibility with contrast stretch')
figure,imhist(K)
```

Morphological operation:

```
BW=im2bw(K,1.0);                % threshold the image (convert the
                                % intensity image to binary image)
figure,imshow(BW),title('threshold image')
seD=strel('diamond',1)          % Create a flat diamond-shape
                                % structuring element
BWerode=imerode(BW,seD);        % Perform erosion to the binary image
figure,imshow(BWerode),title('segmented image')
se90=strel('line',3,90);        % Create a flat, linear structuring
se0=strel('line',3,0);          % element, where 3 is the length,
                                % with 90 and 0 degree angle (in
                                % degrees) of the line respectively,
                                % as measured in a counterclockwise
                                % direction from the horizontal axis
```

```
BWsdil=imdilate(BWerode,[se90 se0]);% perform dilation to the eroded image
figure,imshow(BWsdil),title('dilate gradient mask')
BW2=bwperim(BWsdil);          % find perimeter pixels in the dilated
                              % image (edge detection)
figure,imshow(BW2),title('perimeter determination')
Segout=B;
Segout(BW2)=255;
figure,imshow(Segout),title('outlined image')
```

Feature extraction:

```
[labeled,numObjects]=bwlabel(BWsdil,4);
numObjects                    % determine number of object
data=regionprops(labeled,'all')

[labeled,numObjects2]=bwlabel(BW2,4);     % to measure perimeter
data2=regionprops(labeled,'all')
sumArea=sum([data2.Area])     % total perimeter
```

```
averagePerimeter=sum([data2.Area])/numObjects % average perimeter

for a=1:numObjects
   data(a).Area
   data(a).Eccentricity
   data(a).Orientation
   data(a).Solidity
   circularity=(4*pi*([data(a).Area]))/(averagePerimeter^2)
end
```

To see each of the extracted feature result

```
averageArea=sum([data.Area])/numObjects  % calculating average area
stdConvexArea=std2([data.ConvexArea])    % measure convex area
averageEccentricity=sum([data.Eccentricity])/numObjects % measure average eccentricity
stdEccentricity=std2([data.Eccentricity])
averageOrientation=mean2([data.Orientation]) %measure average orientation
stdOrientation=std2([data.Orientation])
averageSolidity=mean2([data.Solidity])       % measure average solidity
averageCircularity=(4*pi*(averageArea))/(averagePerimeter^2) % measure average circularity
```

# APPENDIX D

# RESULT

## Result of feature extraction

| Samples/features | numObjects | meanPerimeter | meanArea | meanEccentricity | meanOrientation | meanSolidity | meanCircularity |
|---|---|---|---|---|---|---|---|
| **Mammo1** | 37 | 19.514 | 37.1892 | 0.6158 | 36.4627 | 0.9356 | 1.2273 |
| | 75 | 16.667 | 29.9467 | 0.4903 | 48.7045 | 0.9559 | 1.3548 |
| **Mammo2** | (L)71 | 18.085 | 34.9718 | 0.5516 | 49.7319 | 0.9527 | 1.3437 |
| | (R)67 | 22.761 | 46.4925 | 0.6219 | 48.5913 | 0.9313 | 1.1277 |
| **Mammo3** | (L)49 | 23.918 | 47.2857 | 0.5425 | 43.3759 | 0.9401 | 1.0387 |
| | (R)68 | 16.971 | 30.25 | 0.5544 | 43.9824 | 0.9556 | 1.3199 |
| **Mammo4** | (L)77 | 24.429 | 63.7403 | 0.5702 | 41.8898 | 0.9416 | 1.3422 |
| | (R)88 | 22.171 | 47.25 | 0.5653 | 50.5949 | 0.9368 | 1.208 |
| **Mammo5** | (L)108 | 20.361 | 46.9722 | 0.5354 | 41.6824 | 0.9494 | 1.4238 |
| | (R)106 | 18.585 | 40.717 | 0.4626 | 39.4174 | 0.9594 | 1.4814 |
| **Mammo6** | (L)86 | 21 | 47.2093 | 0.5398 | 42.5276 | 0.9484 | 1.3452 |
| | (R)174 | 17.012 | 32.5977 | 0.5249 | 42.7168 | 0.961 | 1.4155 |
| **Mammo7** | (L)77 | 20.286 | 40.6753 | 0.548 | 42.3019 | 0.9397 | 1.2421 |
| | (R)108 | 15.444 | 27.3333 | 0.491 | 32.4537 | 0.9636 | 1.44 |
| **Mammo8** | (L)139 | 12.273 | 18.1223 | 0.3683 | 40.34 | 0.9771 | 1.5118 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (R)153 | 11.294 | 15.6667 | 0.3493 | 37.8013 | 0.9855 | 1.5434 |
| **Mammo9** | (L)135 | 21.378 | 42.6148 | 0.6308 | 34.5104 | 0.9419 | 1.1718 |
| | (R)133 | 20.459 | 41.2256 | 0.5755 | 35.2239 | 0.9424 | 1.2377 |
| **Mammo10** | (L)80 | 30.9 | 93.425 | 0.5802 | 50.1844 | 0.935 | 1.2296 |
| | (R)112 | 23.098 | 58.0625 | 0.5392 | 45.4232 | 0.9496 | 1.3676 |
| **Mammo11** | (L)69 | 22.71 | 51.9275 | 0.5168 | 42.1078 | 0.9425 | 1.2652 |
| | (R)72 | 21.792 | 50.75 | 0.47 | 43.607 | 0.952 | 1.343 |
| **Mammo12** | (L)49 | 23.143 | 51.6939 | 0.5857 | 36.9675 | 0.9218 | 1.2129 |
| | (R)41 | 17.537 | 30.9756 | 0.5051 | 36.4809 | 0.945 | 1.2657 |
| **Mammo13** | 78 | 17.769 | 36.1154 | 0.509 | 43.4528 | 0.9587 | 1.4374 |
| | 85 | 19.106 | 40.1882 | 0.5372 | 46.5265 | 0.954 | 1.3835 |
| **Mammo14** | 178 | 15.865 | 28.3483 | 0.5015 | 40.4568 | 0.9615 | 1.4153 |
| | 127 | 18.016 | 37.6063 | 0.5242 | 42.501 | 0.9622 | 1.456 |
| **Mammo15** | 52 | 23.596 | 51.4038 | 0.5823 | 45.9819 | 0.9205 | 1.1602 |
| | 70 | 18.043 | 35.6286 | 0.5112 | 45.5309 | 0.9456 | 1.3753 |
| **Mammo16** | 94 | 18.128 | 40.7979 | 0.4825 | 38.6445 | 0.9577 | 1.5601 |
| | 47 | 21.596 | 55.1915 | 0.5011 | 48.1825 | 0.9562 | 1.4871 |
| **Mammo17** | 72 | 19.139 | 40.3056 | 0.5768 | 36.4275 | 0.9558 | 1.3827 |
| | 63 | 19.333 | 41.7302 | 0.5012 | 43.0206 | 0.9592 | 1.403 |
| **Mammo18** | 216 | 11.546 | 16.9676 | 0.3349 | 40.5839 | 0.9894 | 1.5994 |
| | 213 | 11.77 | 16.6432 | 0.3523 | 41.559 | 0.9819 | 1.5097 |
| **Mammo19** | 74 | 19.919 | 39.9459 | 0.4885 | 38.6698 | 0.9421 | 1.2652 |
| | 100 | 15.02 | 25.13 | 0.511 | 31.5662 | 0.9619 | 1.3998 |
| **Mammo20** | 45 | 18.356 | 35.5778 | 0.488 | 40.7486 | 0.9491 | 1.3269 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 53 | 16.83 | 38.175 | 0.5895 | 51.4452 | 0.937 | 1.2267 |
| **Mammo21** | 73 | 31.548 | 110.658 | 0.6632 | 53.4332 | 0.9432 | 1.3972 |
| | 110 | 25.082 | 57.3455 | 0.6553 | 37.9914 | 0.9393 | 1.1455 |
| **Mammo22** | 60 | 21.583 | 61.1167 | 0.5413 | 50.3464 | 0.9634 | 1.6487 |
| | 73 | 20 | 40.0548 | 0.5534 | 43.7164 | 0.9531 | 1.2584 |
| **Mammo23** | 148 | 16.595 | 29.3919 | 0.5285 | 38.5567 | 0.9569 | 1.3412 |
| | 103 | 27.767 | 57.9903 | 0.6037 | 35.7579 | 0.9151 | 0.9452 |
| **Mammo24** | 134 | 20.522 | 45.4552 | 0.5543 | 41.3987 | 0.953 | 1.3562 |
| | 139 | 18.086 | 40.8705 | 0.5495 | 37.143 | 0.9591 | 1.5701 |
| **Mammo25** | 157 | 16.045 | 28.7134 | 0.4877 | 40.5464 | 0.9592 | 1.4016 |
| | 90 | 20.478 | 40.6333 | 0.5481 | 37.502 | 0.9426 | 1.2177 |
| **Mammo26** | 144 | 17.542 | 34.7361 | 0.4437 | 37.934 | 0.9565 | 1.4186 |
| | 105 | 18.381 | 36.2952 | 0.5024 | 34.9052 | 0.9549 | 1.35 |
| **Mammo27** | 134 | 18.993 | 37.7388 | 0.521 | 40.6209 | 0.9503 | 1.3147 |
| | 96 | 21.125 | 50.0729 | 0.5158 | 35.1135 | 0.9524 | 1.41 |
| **Mammo28** | 70 | 18.1 | 34.7429 | 0.5398 | 49.1623 | 0.9593 | 1.3327 |
| | 68 | 20.941 | 53.6765 | 0.5491 | 37.2929 | 0.9524 | 1.5381 |
| **Mammo29** | 65 | 32.615 | 93.2769 | 0.6127 | 49.4793 | 0.915 | 1.1019 |
| | 49 | 33.898 | 124.816 | 0.5555 | 49.8664 | 0.9352 | 1.1365 |
| **Mammo30** | 35 | 37.257 | 100.514 | 0.6233 | 49.597 | 0.9203 | 0.91 |
| | 63 | 23.016 | 50.3651 | 0.5009 | 41.9259 | 0.9407 | 1.1948 |
| **Mammo31** | 88 | 18.909 | 43.3523 | 0.5112 | 43.4307 | 0.9614 | 1.5236 |
| | 77 | 20.338 | 45.2078 | 0.5025 | 38.0579 | 0.9444 | 1.3735 |
| **Mammo32** | 98 | 18.245 | 35.7449 | 0.5131 | 40.199 | 0.9524 | 1.3493 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 174 | 13.851 | 21.7816 | 0.4575 | 38.4752 | 0.9698 | 1.4268 |
| **Mammo33** | 57 | 24.316 | 61.8947 | 0.5916 | 50.8033 | 0.9451 | 1.3155 |
| | 50 | 26.88 | 94.68 | 0.4349 | 39.2997 | 0.9385 | 1.6467 |
| **Mammo34** | 49 | 19.082 | 37.2245 | 0.5427 | 48.0256 | 0.9491 | 1.2847 |
| | 49 | 17.918 | 37.6939 | 0.4941 | 40.8166 | 0.9612 | 1.4753 |
| **Mammo35** | 38 | 18.921 | 37.8421 | 0.498 | 34.6626 | 0.9466 | 1.3283 |
| | 60 | 18.617 | 39.6 | 0.4706 | 35.1979 | 0.9526 | 1.4358 |
| **Mammo36** | 77 | 14.883 | 27.6753 | 0.4996 | 43.3572 | 0.9735 | 1.5701 |
| | 60 | 18.15 | 39.2167 | 0.5132 | 52.1013 | 0.9606 | 1.496 |
| **Mammo37** | 129 | 19.574 | 41.6202 | 0.5264 | 35.9957 | 0.9558 | 1.3651 |
| | 129 | 19.147 | 38.9535 | 0.5629 | 36.2134 | 0.9526 | 1.3352 |
| **Mammo38** | 60 | 20.2 | 43.2 | 0.5256 | 40.9849 | 0.9396 | 1.3304 |
| | 72 | 18.75 | 39.625 | 0.482 | 40.311 | 0.9547 | 1.4164 |
| **Mammo39** | 82 | 16.915 | 30.4634 | 0.5364 | 42.2533 | 0.956 | 1.338 |
| | 68 | 21.588 | 46.9853 | 0.5546 | 46.3599 | 0.942 | 1.2669 |
| **Mammo40** | 94 | 19.628 | 44.1277 | 0.524 | 43.7941 | 0.9623 | 1.4394 |
| | 91 | 21.429 | 47.2198 | 0.6102 | 44.1867 | 0.9485 | 1.2923 |
| **Mammo41** | 21 | 36.429 | 117.286 | 0.531 | 59.1507 | 0.9472 | 1.1106 |
| | 40 | 24.35 | 61.875 | 0.5807 | 55.4068 | 0.9474 | 1.3114 |
| **Mammo42** | 137 | 15.569 | 28.562 | 0.4942 | 39.7061 | 0.973 | 1.4807 |
| | 133 | 16.361 | 30.0451 | 0.5196 | 44.3736 | 0.9622 | 1.4105 |
| **Mammo43** | 62 | 23.855 | 59.7581 | 0.5725 | 40.661 | 0.9427 | 1.3196 |
| | 82 | 23.146 | 48.9024 | 0.5764 | 45.5499 | 0.9524 | 1.147 |
| **Mammo44** | 46 | 26.196 | 77.2609 | 0.5752 | 43.7492 | 0.9464 | 1.4149 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 59 | 20.102 | 54.0169 | 0.3529 | 43.3191 | 0.9706 | 1.6799 |
| **Mammo45** | 44 | 30.296 | 82.7955 | 0.6855 | 43.1782 | 0.8974 | 1.1336 |
| | 49 | 22.531 | 50.5306 | 0.5885 | 35.4547 | 0.9347 | 1.2509 |
| **Mammo46** | 57 | 17.404 | 33.2632 | 0.4315 | 45.2226 | 0.9356 | 1.3801 |
| | 61 | 16.393 | 30.9016 | 0.4472 | 43.9282 | 0.9559 | 1.4449 |
| **Mammo47** | 39 | 21.718 | 42.5641 | 0.3653 | 49.3794 | 0.9527 | 1.134 |
| | 62 | 18.048 | 34.9839 | 0.4787 | 41.9559 | 0.9313 | 1.3496 |
| **Mammo48** | 110 | 19.327 | 40.1364 | 0.5537 | 50.6007 | 0.9401 | 1.3502 |
| | 92 | 20.196 | 40.163 | 0.5821 | 52.5643 | 0.9556 | 1.2374 |
| **Mammo49** | 63 | 16.333 | 30.8889 | 0.4565 | 46.2244 | 0.9416 | 1.455 |
| | 87 | 16.345 | 31.023 | 0.4753 | 49.4027 | 0.9368 | 1.4592 |
| **Mammo50** | 45 | 17.467 | 32.8889 | 0.513 | 47.9979 | 0.9494 | 1.3547 |
| | 55 | 19.146 | 38.5091 | 0.5455 | 50.9736 | 0.9594 | 0.3202 |
| **Mammo51** | 59 | 16.119 | 27.4746 | 0.4664 | 35.8841 | 0.9484 | 1.3289 |
| | 91 | 18.868 | 40.3077 | 0.5157 | 36.3498 | 0.961 | 1.4228 |
| **Mammo52** | 81 | 23 | 52.7284 | 0.5527 | 43.0157 | 0.9397 | 1.2526 |
| | 116 | 18.328 | 36.069 | 0.5221 | 39.6452 | 0.9636 | 1.3494 |
| **Mammo53** | 100 | 20.93 | 45.71 | 0.5476 | 47.6376 | 0.9771 | 1.3112 |
| | 33 | 36.182 | 181.939 | 0.5383 | 48.4154 | 0.9855 | 1.7465 |
| **Mammo54** | 41 | 15.146 | 25.5122 | 0.5612 | 35.8244 | 0.9419 | 1.3975 |
| | 164 | 14.305 | 24.9268 | 0.4289 | 34.3457 | 0.9424 | 1.5308 |
| **Mammo55** | 133 | 16.188 | 29.9398 | 0.474 | 47.4555 | 0.935 | 1.4357 |
| | 150 | 16.447 | 27.6467 | 0.5171 | 50.0849 | 0.9496 | 1.2844 |
| **Mammo56** | 103 | 19.252 | 38.1748 | 0.5174 | 48.6881 | 0.9425 | 1.2942 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 117 | 18.897 | 37.4359 | 0.5025 | 47.8094 | 0.952 | 1.3173 |
| **Mammo57** | 123 | 18.431 | 35.8293 | 0.5681 | 42.5961 | 0.9218 | 1.3254 |
| | 223 | 18.014 | 35.2691 | 0.484 | 39.9691 | 0.945 | 1.3659 |
| **Mammo58** | 58 | 18.414 | 39.3276 | 0.559 | 35.8162 | 0.9587 | 1.4615 |
| | 159 | 18.459 | 39.6289 | 0.5086 | 38.8172 | 0.954 | 1.4575 |
| **Mammo59** | 20 | 19.95 | 41.4 | 0.4818 | 29.9521 | 0.9615 | 1.3071 |
| | 84 | 19.179 | 41.3452 | 0.5761 | 46.711 | 0.9622 | 1.4125 |
| **Mammo60** | 468 | 14.38 | 25.2457 | 0.3699 | 36.2931 | 0.9205 | 1.5341 |
| | 279 | 15.423 | 27.0036 | 0.4501 | 42.6219 | 0.9456 | 1.4266 |
| **Mammo61** | 143 | 18.629 | 37.4126 | 0.5457 | 48.0879 | 0.9577 | 1.3547 |
| | 124 | 19.145 | 36.9032 | 0.5829 | 51.1581 | 0.9562 | 1.2652 |
| **Mammo62** | 162 | 17.648 | 32.9753 | 0.5184 | 32.2615 | 0.9558 | 1.3305 |
| | 178 | 14.674 | 25.5281 | 0.4431 | 32.4618 | 0.9592 | 1.4898 |
| **Mammo63** | 111 | 16.252 | 33.9431 | 0.4874 | 42.7645 | 0.9894 | 1.3345 |
| | 123 | 17.878 | 30.7297 | 0.4317 | 37.3941 | 0.9819 | 1.462 |
| **Mammo64** | 371 | 14.973 | 27.3181 | 0.3998 | 36.623 | 0.9421 | 1.5312 |
| | 345 | 12.849 | 19.5768 | 0.38 | 33.9609 | 0.9619 | 1.49 |
| **Mammo65** | 124 | 15.927 | 29.4194 | 0.5038 | 37.8724 | 0.9491 | 1.4573 |
| | 554 | 14.731 | 25.5162 | 0.398 | 34.6575 | 0.937 | 1.4776 |
| **Mammo66** | 59 | 16.712 | 30.8814 | 0.4626 | 47.2685 | 0.9432 | 1.3895 |
| | 62 | 16.081 | 29.4032 | 0.4756 | 48.1037 | 0.9393 | 1.429 |
| **Mammo67** | 235 | 16.979 | 32.0936 | 0.4661 | 41.3853 | 0.9634 | 1.399 |
| | 352 | 16.574 | 32.7159 | 0.4432 | 39.9467 | 0.9531 | 1.4966 |

**Result of preliminary data processing**

Random number generation and actual output:

| | | *eccentricity* | *orientation* | *circularity* | *area* | **random number** | **output** |
|---|---|---|---|---|---|---|---|
| | *R* | 0.6158 | 36.4627 | 1.2273 | 37.1892 | 17.69286 | 1 |
| **mammo1** | *L* | 0.4903 | 48.7045 | 1.3548 | 29.9467 | 72.45753 | 0 |
| | *R* | 0.6219 | 48.5913 | 1.1277 | 46.4925 | 92.43458 | 0 |
| **mammo2** | *L* | 0.5516 | 49.7319 | 1.3437 | 34.9718 | 45.23835 | 1 |
| | *R* | 0.5544 | 43.9824 | 1.3199 | 30.25 | 85.52177 | 1 |
| **mammo3** | *L* | 0.5425 | 43.3759 | 1.0387 | 47.2857 | 12.33604 | 1 |
| | *R* | 0.5653 | 50.5949 | 1.208 | 47.25 | 16.15198 | 1 |
| **mammo4** | *L* | 0.5702 | 41.8898 | 1.3422 | 63.7403 | 28.67238 | 0 |
| | *R* | 0.4626 | 39.4174 | 1.4814 | 40.717 | 56.43843 | 0 |
| **mammo5** | *L* | 0.5354 | 41.6824 | 1.4238 | 46.9722 | 8.946104 | 1 |
| | *R* | 0.5249 | 42.7168 | 1.4155 | 32.5977 | 91.14145 | 0 |
| **mammo6** | *L* | 0.5398 | 42.5276 | 1.3452 | 47.2093 | 32.50645 | 1 |
| | *R* | 0.491 | 32.4537 | 1.44 | 27.3333 | 68.77151 | 0 |
| **mammo7** | *L* | 0.548 | 42.3019 | 1.2421 | 40.6753 | 46.84872 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *R* | 0.3493 | 37.8013 | 1.5434 | 15.6667 | 6.640828 | 1 |
| **mammo8** | *L* | 0.3683 | 40.34 | 1.5118 | 18.1223 | 30.76012 | 1 |
| | *R* | 0.5755 | 35.2239 | 1.2377 | 41.2256 | 55.65893 | 1 |
| **mammo9** | *L* | 0.6308 | 34.5104 | 1.1718 | 42.6148 | 18.43611 | 0 |
| | *R* | 0.5392 | 45.4232 | 1.3676 | 58.0625 | 69.44527 | 1 |
| **mammo10** | *L* | 0.5802 | 50.1844 | 1.2296 | 93.425 | 91.57048 | 1 |
| | *R* | 0.47 | 43.607 | 1.343 | 50.75 | 64.70479 | 0 |
| **mammo11** | *L* | 0.5168 | 42.1078 | 1.2652 | 51.9275 | 64.43892 | 0 |
| | *R* | 0.5051 | 36.4809 | 1.2657 | 30.9756 | 46.35624 | 1 |
| **mammo12** | *L* | 0.5857 | 36.9675 | 1.2129 | 51.6939 | 1.151067 | 0 |
| | *R* | 0.509 | 43.4528 | 1.4374 | 36.1154 | 61.901 | 0 |
| **mammo13** | *L* | 0.5372 | 46.5265 | 1.3835 | 40.1882 | 85.42811 | 1 |
| | *R* | 0.5015 | 40.4568 | 1.4153 | 28.3483 | 97.44395 | 1 |
| **mammo14** | *L* | 0.5242 | 42.501 | 1.456 | 37.6063 | 40.39213 | 0 |
| | *R* | 0.5823 | 45.9819 | 1.1602 | 51.4038 | 39.75161 | 1 |
| **mammo15** | *L* | 0.5112 | 45.5309 | 1.3753 | 35.6286 | 23.89868 | 0 |
| | *R* | 0.4825 | 38.6445 | 1.5601 | 40.7979 | 90.70336 | 1 |
| **mammo16** | *L* | 0.5011 | 43.0206 | 1.4871 | 55.1915 | 83.28599 | 1 |
| **mammo17** | *R* | 0.5768 | 36.4275 | 1.3827 | 40.3056 | 11.84658 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *L* | 0.5012 | 41.559 | 1.403 | 41.7302 | 90.8786 | 1 |
| | *R* | 0.3349 | 40.5839 | 1.5994 | 16.9676 | 73.1887 | 1 |
| **mammo18** | *L* | 0.3523 | 31.5662 | 1.5097 | 16.6432 | 76.89889 | 1 |
| | *R* | 0.4885 | 38.6698 | 1.2652 | 39.9459 | 20.82598 | 0 |
| **mammo19** | *L* | 0.511 | 51.4452 | 1.3998 | 25.13 | 61.43873 | 1 |
| | *R* | 0.488 | 40.7486 | 1.3269 | 35.5778 | 70.21268 | 1 |
| **mammo20** | *L* | 0.5895 | 37.9914 | 1.2267 | 38.175 | 62.12458 | 1 |
| | *R* | 0.6632 | 53.4332 | 1.3972 | 110.658 | 38.98419 | 1 |
| **mammo21** | *L* | 0.6553 | 43.7164 | 1.1455 | 57.3455 | 80.2284 | 1 |
| | *R* | 0.5413 | 50.3464 | 1.6487 | 61.1167 | 7.728507 | 1 |
| **mammo22** | *L* | 0.5534 | 35.7579 | 1.2584 | 40.0548 | 22.93487 | 1 |
| | *R* | 0.5285 | 38.5567 | 1.3412 | 29.3919 | 86.43422 | 1 |
| **mammo23** | *L* | 0.6037 | 37.143 | 0.9452 | 57.9903 | 22.36686 | 0 |
| | *R* | 0.5543 | 41.3987 | 1.3562 | 45.4552 | 8.477798 | 1 |
| **mammo24** | *L* | 0.5495 | 37.502 | 1.5701 | 40.8705 | 99.31114 | 0 |
| | *R* | 0.4877 | 40.5464 | 1.4016 | 28.7134 | 48.7431 | 0 |
| **mammo25** | *L* | 0.5481 | 34.9052 | 1.2177 | 40.6333 | 74.97732 | 0 |
| | *R* | 0.4437 | 37.934 | 1.4186 | 34.7361 | 65.62328 | 1 |
| **mammo26** | *L* | 0.5024 | 35.1136 | 1.35 | 36.2952 | 23.64791 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *R* | 0.521 | 40.6209 | 1.3147 | 37.7388 | 94.15674 | 1 |
| **mammo27** | *L* | 0.5158 | 37.2929 | 1.41 | 50.0729 | 3.979034 | 1 |
| | *R* | 0.5398 | 49.1623 | 1.3327 | 34.7429 | 94.67641 | 0 |
| **mammo28** | *L* | 0.5491 | 49.8664 | 1.5381 | 53.6765 | 56.90976 | 0 |
| | *R* | 0.6127 | 49.4793 | 1.1019 | 93.2769 | 36.27103 | 1 |
| **mammo29** | *L* | 0.5555 | 41.9259 | 1.365 | 124.816 | 30.4731 | 1 |
| | *R* | 0.6233 | 49.597 | 0.91 | 100.514 | 26.068 | 1 |
| **mammo30** | *L* | 0.5009 | 38.0579 | 1.1948 | 50.3651 | 36.97198 | 1 |
| | *R* | 0.5112 | 43.4307 | 1.5236 | 43.3523 | 65.87707 | 0 |
| **mammo31** | *L* | 0.5025 | 38.4752 | 1.3735 | 45.2078 | 75.79913 | 0 |
| | *R* | 0.5131 | 40.199 | 1.3493 | 35.7449 | 50.17823 | 0 |
| **mammo32** | *L* | 0.4575 | 38.4752 | 1.4268 | 21.7816 | 73.95007 | 1 |
| | *R* | 0.5916 | 50.8033 | 1.3155 | 61.8947 | 36.36772 | 1 |
| **mammo33** | *L* | 0.4349 | 39.2997 | 1.6467 | 94.68 | 78.02887 | 1 |
| | *R* | 0.5427 | 48.0256 | 1.2847 | 37.2245 | 26.43962 | 1 |
| **mammo34** | *L* | 0.4941 | 40.8166 | 1.4753 | 37.6939 | 34.06244 | 1 |
| | *R* | 0.498 | 34.6626 | 1.3283 | 37.8421 | 84.99908 | 0 |
| **mammo35** | *L* | 0.4706 | 35.1979 | 1.4358 | 39.6 | 71.1644 | 1 |
| **mammo36** | *R* | 0.4996 | 43.3572 | 1.5701 | 27.6753 | 50.14197 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *L* | 0.5132 | 52.1013 | 1.496 | 39.2167 | 79.1377 | 1 |
| | *R* | 0.5264 | 35.9957 | 1.3651 | 41.6202 | 16.7774 | 1 |
| **mammo37** | *L* | 0.5629 | 36.2134 | 1.3352 | 38.9535 | 16.23054 | 1 |
| | *R* | 0.5256 | 40.9849 | 1.3304 | 43.2 | 80.455 | 0 |
| **mammo38** | *L* | 0.482 | 40.311 | 1.4164 | 39.625 | 62.34513 | 0 |
| | *R* | 0.5364 | 42.2533 | 1.338 | 30.4634 | 40.57945 | 1 |
| **mammo39** | *L* | 0.5546 | 46.3599 | 1.2669 | 46.9853 | 30.9263 | 0 |
| | *R* | 0.524 | 43.7941 | 1.4394 | 44.1277 | 55.93689 | 1 |
| **mammo40** | *L* | 0.6102 | 44.1867 | 1.2923 | 47.2198 | 35.61843 | 1 |
| | *R* | 0.531 | 59.1507 | 1.1106 | 117.286 | 51.32633 | 1 |
| **mammo41** | *L* | 0.5807 | 55.4068 | 1.3114 | 61.875 | 30.83868 | 1 |
| | *R* | 0.4942 | 39.7061 | 1.4807 | 28.562 | 92.20194 | 0 |
| **mammo42** | *L* | 0.5196 | 44.3736 | 1.4105 | 30.0451 | 15.48125 | 1 |
| | *R* | 0.5725 | 40.661 | 1.3196 | 59.7581 | 47.82461 | 1 |
| **mammo43** | *L* | 0.5764 | 45.5499 | 1.147 | 48.9024 | 22.88049 | 0 |
| | *R* | 0.5752 | 43.7492 | 1.4149 | 77.2609 | 9.988464 | 1 |
| **mammmo44** | *L* | 0.3529 | 43.3191 | 1.6799 | 54.0169 | 74.72958 | 1 |
| | *R* | 0.6855 | 43.1782 | 1.1336 | 82.7955 | 68.14307 | 0 |
| **mammo45** | *L* | 0.5885 | 35.4547 | 1.2509 | 50.5306 | 64.03406 | 0 |

Segmentation of data:

| all data | training | validation | testing |
|---|---|---|---|
| 17.69286 | 94.67641 | 22.93487 | 91.57048 |
| 72.45753 | 36.36772 | 55.93689 | 26.068 |
| 92.43458 | 83.28599 | 8.946104 | 73.95007 |
| 45.23835 | 61.901 | 16.7774 | 47.82461 |
| 85.52177 | 46.84872 | 83.28599 | 68.77151 |
| 12.33604 | 79.1377 | 35.61843 | 15.48125 |
| 16.15198 | 45.23835 | 70.21268 | 8.477798 |
| 28.67238 | 92.20194 | 62.12458 | 91.14145 |
| 56.43843 | 18.43611 | 74.97732 | 50.14197 |
| 8.946104 | 28.67238 | 36.27103 | 71.1644 |
| 91.14145 | 6.640828 | 56.90976 | 9.988464 |
| 32.50645 | 85.52177 | 47.82461 | 30.83868 |
| 68.77151 | 32.50645 | 97.44395 | 90.8786 |
| 46.84872 | 65.87707 | 80.2284 | |
| 6.640828 | 22.88049 | 61.43873 | |
| 30.76012 | 68.14307 | 65.87707 | |

| | | |
|---|---|---|
| 55.65893 | 8.477798 | 46.35624 |
| 18.43611 | 9.988464 | 30.83868 |
| 69.44527 | 23.64791 | 46.84872 |
| 91.57048 | 70.21268 | 64.70479 |
| 64.70479 | 50.17823 | 76.89889 |
| 64.43892 | 26.068 | 30.4731 |
| 46.35624 | 39.75161 | 39.75161 |
| 1.151067 | 86.43422 | 91.14145 |
| 61.901 | 30.76012 | 85.52177 |
| 85.42811 | 69.44527 | 94.67641 |
| 97.44395 | 8.946104 | 26.43962 |
| 40.39213 | 7.728507 | 65.62328 |
| 39.75161 | 16.7774 | 30.9263 |
| 23.89868 | 68.77151 | 48.7431 |
| 90.70336 | 78.02887 | 92.20194 |
| 83.28599 | 80.455 | 9.988464 |
| 11.84658 | 16.15198 | 91.57048 |
| 90.8786 | 74.97732 | 3.979034 |
| 73.1887 | 1.151067 | 68.14307 |

| | | |
|---|---|---|
| 76.89889 | 64.43892 | 78.02887 |
| 20.82598 | 30.4731 | 68.77151 |
| 61.43873 | 17.69286 | 56.43843 |
| 70.21268 | 56.43843 | |
| 62.12458 | | |
| 38.98419 | | |
| 80.2284 | | |
| 7.728507 | | |
| 22.93487 | | |
| 86.43422 | | |
| 22.36686 | | |
| 8.477798 | | |
| 99.31114 | | |
| 48.7431 | | |
| 74.97732 | | |
| 65.62328 | | |
| 23.64791 | | |
| 94.15674 | | |
| 3.979034 | | |

94.67641

56.90976

36.27103

30.4731

26.068

36.97198

65.87707

75.79913

50.17823

73.95007

36.36772

78.02887

26.43962

34.06244

84.99908

71.1644

50.14197

79.1377

16.7774

16.23054

80.455

62.34513

40.57945

30.9263

55.93689

35.61843

51.32633

30.83868

92.20194

15.48125

47.82461

22.88049

9.988464

74.72958

68.14307

64.03406

ANOVA test:

Anova: Single Factor

SUMMARY

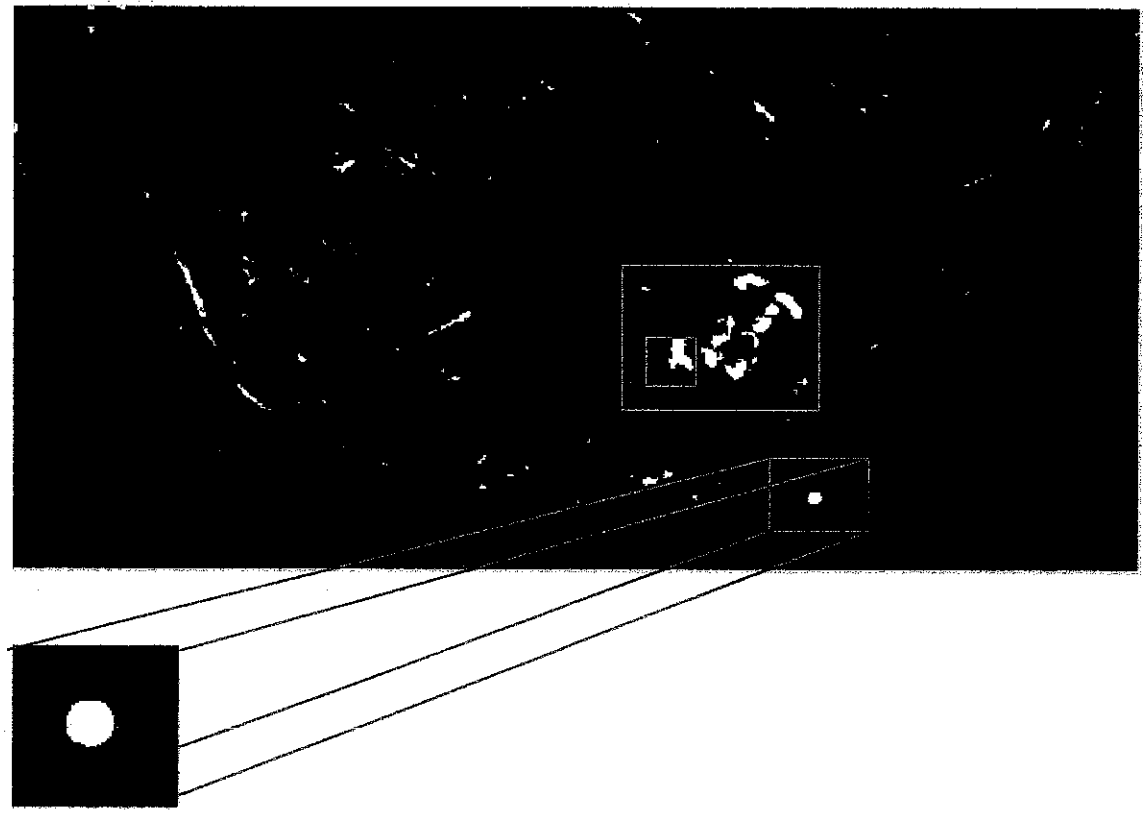
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
|---|---|---|---|---|
| all data | 90 | 4615.968 | 51.28853 | 768.9651 |
| training | 39 | 1805.331 | 46.29055 | 827.0245 |
| validation | 38 | 2124.874 | 55.91773 | 680.459 |
| testing | 13 | 676.2973 | 52.02287 | 989.4755 |

ANOVA

| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
|---|---|---|---|---|---|---|
| Between Groups | 1795.045 | 3 | 598.3484 | 0.769155 | 0.51272 | 2.655939 |
| Within Groups | 136915.5 | 176 | 777.9291 | | | |
| Total | 138710.6 | 179 | | | | |

# APPENDIX E
# VERIFICATION

Cropped area from mammo20-1



numObjects = 1
sumArea = 442
meanArea = 442
meanEccentricity = 0.1412
meanOrientation = 17.6378
meanSolidity = 0.9736
meanEquivDiameter = 23.7228

Perimeter:

sumArea = 68
meanperimeter = 68

>> circularity = (4*pi*([meanArea]))/(meanperimeter^2) = 1.2012
>> circularity = (meanperimeter^2)/(4*pi*([meanArea])) = 0.8325

numObjects = 1
sumArea = 1387
meanArea = 1387
meanEccentricity = 0.7906
meanOrientation = -75.7992
meanSolidity = 0.7642
meanEquivDiameter = 42.0236

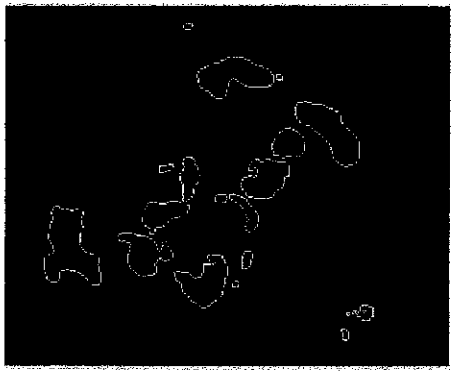
Perimeter:

sumArea = 173
meanperimeter = 173

>> circularity = (4*pi*([meanArea]))/(meanperimeter^2) = 0.5824
>> circularity = (meanperimeter^2)/(4*pi*([meanArea])) = 1.7171

numObjects = 18
sumArea = 7575
meanArea = 420.8333
meanEccentricity = 0.6822
meanOrientation = -1.8235
meanSolidity = 0.8921
meanEquivDiameter = 19.2224

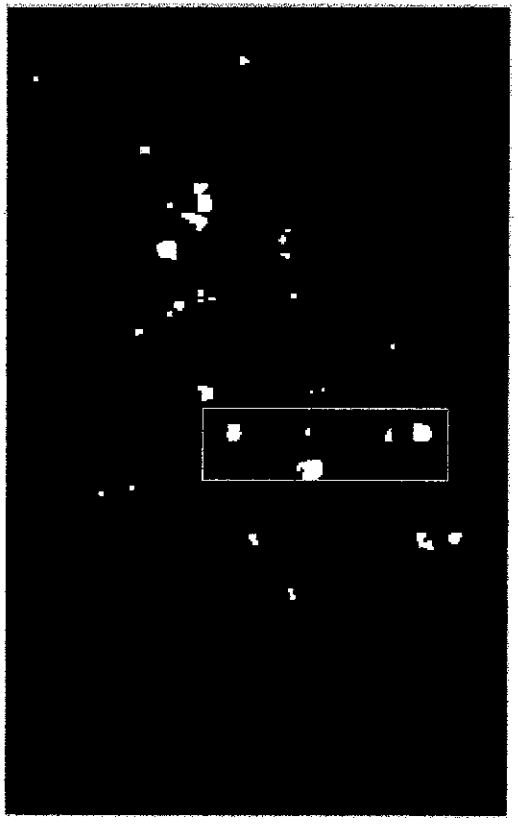
Perimeter:



sumArea = 1269
meanperimeter = 70.5000

>> circularity = (4*pi*([meanArea]))/(meanperimeter^2) = 1.0640
>> circularity = (meanperimeter^2)/(4*pi*([meanArea])) = 0.9398

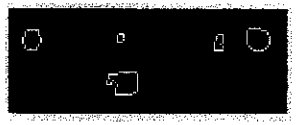
Cropped area from case21_Lcc



numObjects = 5
sumArea = 424

meanArea = 84.8000
meanEccentricity = 0.6332
meanOrientation = 35.6147
meanSolidity = 0.9578
meanEquivDiameter = 9.8129

Perimeter:



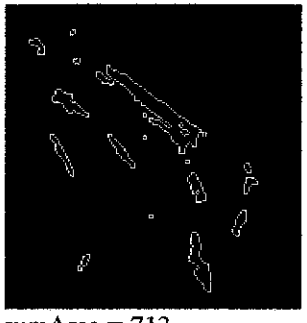
sumArea = 148
meanperimeter = 29.6000

>> circularity = (4*pi*([meanArea]))/(meanperimeter^2) = 1.2162
>> circularity = (meanperimeter^2)/(4*pi*([meanArea])) = 0.8222

Cropped area from case30_Lcc

numObjects = 17
sumArea = 1946
meanArea = 114.4706
meanEccentricity = 0.6718
meanOrientation = -13.3507
meanSolidity = 0.8841
meanEquivDiameter = 9.7684

Perimeter:



sumArea = 713
meanperimeter = 41.9412

>> circularity = (4*pi*([meanArea]))/(meanperimeter^2) = 0.8178
>> circularity = (meanperimeter^2)/(4*pi*([meanArea])) = 1.2229

# APPENDIX F

# SAMPLE DETAILS

| Samples | Description | Further evaluation needed? |
|---|---|---|
| **Mammo1** | - asymmetrical density lesion (R) | yes |
| **Mammo2** | - rim/egg-shelled calcification (L) | no |
| **Mammo3** | - vascular calcification (R)<br>- vascular calcification (L) | no |
| **Mammo4** | - asymmetrical density lesion (R) | yes |
| **Mammo5** | - architectural distortion lesions (L) | yes |
| **Mammo6** | - round/oval lesions with obscured margins (L) | yes |
| **Mammo7** | - spherical/lucent calcification (R)<br>- round and spherical/lucent calcification (L) | no |
| **Mammo8** | - round/oval lesions with speculated margins (L) | yes |
| **Mammo9** | - round calcification (R)<br>- round calcification (L) | no |
| **Mammo10** | - round/oval lesions with circumscribed (R) | yes |
| **Mammo11** | - spherical/lucent calcification (R)<br>- spherical/lucent and rim/egg-shelled calcification (L) | no |
| **Mammo12** | No suspicious lesion (normal) | no |
| **Mammo13** | - Pleomorphic/heterogeneous calcifications (R) | yes |
| **Mammo14** | - Asymmetrical density lesions (L) | yes |
| **Mammo15** | - irregular lesions with speculated margins and pleomorphic/heterogeneous calcifications (R) | yes |
| **Mammo16** | - round/oval lesions with circumscribed margins (R) | yes |
| **Mammo17** | - spherical/lucent calcifications (R)<br>- round/oval lesions with speculated margins and spherical/lucent calcifications (L) | yes |
| **Mammo18** | - round and punctuate calcifications (R)<br>- round and punctuate calcifications (L) | no |
| **Mammo19** | - round/oval lesions with ill-defined margins (L) | yes |

| | | |
|---|---|---|
| **Mammo20** | - vascular calcifications (R)<br>- vascular calcifications and pleomorphic/heterogeneous calcifications (L) | yes |
| **Mammo21** | - Rod-shaped calcifications (R)<br>- Rod-shaped calcifications (L) | no |
| **Mammo22** | - round calcifications (R)<br>- round/oval lesion with obscured margins and spherical/lucent calcification (L) | yes |
| **Mammo23** | - pleomorphic/heterogeneous and fine linear branching calcifications (R) | yes |
| **Mammo24** | - irregular lesions with obscured margins (R) | yes |
| **Mammo25** | no suspicious lesion (normal) | no |
| **Mammo26** | - round/oval lesions with circumscribed margins (R)<br>- irregular lesions with speculated margins | yes |
| **Mammo27** | - vascular calcifications (R)<br>- asymmetrical density lesions and vascular calcifications (L) | yes |
| **Mammo28** | no suspicious lesion (normal) | no |
| **Mammo29** | - Round/oval lesions with circumscribed margins and coarse/popcorn calcifications (R)<br>- Round/oval lesions with circumscribed margins (L) | yes |
| **Mammo30** | - vascular calcifications (R)<br>- vascular calcifications (L) | no |
| **Mammo31** | no suspicious lesion (normal) | no |
| **Mammo32** | - punctuate and indistinct/amorphous calcifications (L) | yes |
| **Mammo33** | - rim/egg-shelled calcification (R)<br>- rim/egg-shelled calcification (L) | no |
| **Mammo34** | - Round/oval lesions with circumscribed margins (R)<br>- Round/oval lesions with circumscribed margins (L) | yes |
| **Mammo35** | - pleomorphic/heterogeneous calcifications (L) | yes |
| **Mammo36** | - round/oval lesions with circumscribed margins (L) | no |
| **Mammo37** | - irregular lesions with obscured margins (L) | yes |
| **Mammo38** | - round/oval lesions with circumscribed margins and round calcifications (R)<br>- round/oval and irregular lesions with circumscribed and speculated margins (L) | yes |
| **Mammo39** | No suspicious lesion (normal) | no |
| **Mammo40** | - asymmetric density and pleomorphic/heterogeneous calcification (R) | yes |
| **Mammo41** | - rod-shaped and round calcification (R)<br>- rod-shaped and round calcification (L) | no |

| | | |
|---|---|---|
| **Mammo42** | - spherical/lucent calcification (R)<br>- spherical/lucent calcification (L) | no |
| **Mammo43** | - asymmetrical density lesions (L) | yes |
| **Mammo44** | - round/oval and irregular lesions with circumscribed (R) | no |
| **Mammo45** | - dystrophic calcifications (R)<br>- dystrophic calcifications (L) | no |
| **Mammo46** | Bilateral calcification | benign |
| **Mammo47** | Large dense calcification (R) | benign |
| **Mammo48** | Calcification (R & L) | benign |
| **Mammo49** | Scattered small calcifications | benign |
| **Mammo50** | Microcalcification (R) | |
| **Mammo51** | - auxiliary lymph nodes (R)<br>- small round calcification (L) | benign |
| **Mammo52** | Bilateral dense breast | |
| **Mammo53** | Popcorn calcified lesion (L) | |
| **Mammo54** | normal | no |
| **Mammo55** | Mass (R) | no |
| **Mammo56** | Dense breast | no |
| **Mammo57** | Normal | no |
| **Mammo58** | Normal | no |
| **Mammo59** | - Coarse calcification (R)<br>- irregular density (L) | |
| **Mammo60** | Irregular shape density and small breast cyst | |
| **Mammo61** | Heterogeneous dense breast and small cyst (R & L) | |
| **Mammo62** | - large density lesion/lesion/calcification (R)<br>- suspicious lesion (L) | |
| **Mammo63** | Macrocalcification and thickening | |
| **Mammo64** | Coarse and small calcification (R) | benign |
| **Mammo65** | Calcification | benign |
| **Mammo66** | Dense breast cyst | |
| **Mammo67** | Coarse calcification | benign |