THE APPLICATION OF IMAGE PROCESSING AND MACHINE LEARNING TECHNIQUES FOR DETECTION AND CLASSIFICATION OF CANCEROUS TISSUES IN DIGITAL MAMMOGRAMS

JAWAD NAGI

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2011

THE APPLICATION OF IMAGE PROCESSING AND MACHINE LEARNING TECHNIQUES FOR DETECTION AND CLASSIFICATION OF CANCEROUS TISSUES IN DIGITAL MAMMOGRAMS

BY

JAWAD NAGI

DISSERTATION SUBMITTED IN FULFILMENT

OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF COMPUTER SCIENCE

FACULTY OF COMPUTER SCIENCE

AND INFORMATION TECHNOLOGY

UNIVERSITY OF MALAYA

KUALA LUMPUR

2011

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **Jawad Nagi** (I.C./Passport No: AD9990581)

Registration/Matric No: **WGA080040**

Name of Degree: **Masters of Computer Science**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

The Application of Image Processing and Machine Learning Techniques for Detection and Classification of Cancerous Tissues in Digital Mammograms

Field of Study: **Medical Informatics**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:
Designation:

ABSTRACT

Breast cancer is one of the most common kinds of cancer, as well as the leading cause of mortality among women. Mammography is currently the most effective imaging modality for the detection of breast cancer and the diagnosis of the anomalies which can identify cancerous cells. Retrospective studies show that, in current breast cancer screenings approximately 15 to 30 percent of breast cancer cases are missed by radiologists. With the advances in digital image processing techniques, it is envisaged that radiologists will have opportunities to decrease this margin of error and hence, improve their diagnosis.

Digital mammograms have become the most effective techniques for the detection of breast cancer. The goal of this research is to increase the diagnostic accuracy of image processing and machine learning techniques for optimum classification between malignant and benign abnormalities in digital mammograms by reducing the number of misclassified cancers. In this research, digital mammography images are obtained from Malaysian patients who are treated at the University of Malaya Medical Centre (UMMC) from 2008 to 2010. This database consists of standard images of dense, fatty and fatty-glandular breasts, which are classified into three categories: normal, benign and malignant, using the results obtained from biopsies.

Image processing techniques are applied in this research to enhance the mammogram images for the computerized detection of breast cancer. Image processing algorithms used for mammogram image processing include morphological operations and thresholding techniques. As the pectoral muscle in digital mammograms can bias the detection results, it should be suppressed from

the mammograms. This research employs a seeded region growing technique for the segmenting the breast tissue from the pectoral muscle.

Malignant and benign abnormalities are selected from the segmented images using the Ground Truth (GT) data and markings obtained from the radiologists' interpretation of the mammography datasets, which correspond to the Regions of Interest (ROIs) or abnormal regions (samples). Texture based features are extracted from the ROI samples using Gray Level Co-Occurrence Matrices (GLCMs). For the purpose of pattern classification between malignant and benign samples, the optimum subset of texture features are modeled using a Support Vector Machine (SVM). The SVM is trained using two-thirds of the total samples where the remaining one-third of samples are used for testing and validation. The binary classification accuracy of the developed system is measured using the Receiver Operating Characteristic (ROC) analysis with performance measures such as sensitivity, specificity and the Area Under the Curve (AUC). To perform a comparative study, machine learning algorithms other than the SVM, namely, Artificial Neural Networks (ANNs) are evaluated in this research.

The experimental results obtained from the system developed in this research prove to be beneficial for the automated detection of breast cancer. The proposed technique will improve the diagnostic accuracy and consistency of the radiologists' image interpretation in the diagnosis of breast cancer. The resulting computerized breast cancer detection system will subsequently act as a second reader after the manual detection by the radiologist and it is believed that this would aid the radiologist in the mammogram screening process.

ACKNOWLEDGEMENTS

First and foremost, I wish to thank God for giving me strength and courage to complete this thesis and research, and also to those who have assisted and inspired me throughout this research.

There are so many people to whom I am indebted for their assistance during my endeavors to complete my Masters candidature in Computer Science at University of Malaya (UM). First and foremost, I would like to express my gratitude to my supervisor Assoc. Prof. Datin Dr. Sameem Abdul Kareem from University of Malaya, Malaysia, whose invaluable guidance and support was very helpful throughout my research. A similar level of gratitude is due to Dr. Farrukh Hafiz Nagi and Mr. Syed Khaleel Ahmed from Universiti Tenaga Nasional, Malaysia. It is unlikely that I would have reached completion without their encouragement and support.

I express my appreciation to everyone involved directly and indirectly to the success of this research. Last but not least, my family for their understanding, support, patience, and encouragement. Thank you for all the support, comments and guidance.

DEDICATION

This thesis is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

TABLE OF CONTENTS

CHAPTER 7 – CONCLUSION AND FUTURE WORK

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACR	American Cancer Society
ALOE	Analysis of Local Oriented Edges
ANN	Artificial Neural Network
ART	Adaptive Resonance Theory
ASR	Age-Standardized Rate
AUC	Area Under Curve
BI-RADS	Breast Imaging Reporting and Data System
BP	Back-Propagation
BPNN	Back-Propagation Neural Network
CM	Completeness
CNN	Convolutional Neural Network
CR	Correctness
CV	Cross-Validation
DCIS	Ductal Carcinoma In Situ
DDSM	Digital Database for Screening Mammography
DoG	Difference of Gaussian
ELM	Extreme Learning Machine
EU25	European Union
FFDM	Full Field Digital Mammography
FIS	Fuzzy Inference System
FN	False Negative
FP	False Positive
FPF	False Positive Fraction
GA	Genetic Algorithm
GLCM	Gray Level Co-occurrence Matrix
GLDM	Gray Level Difference Method
GLRLM	Gray Level Run Length Matrix
GT	Ground Truth
HIP	Health Insurance Plan
KA	Kernel Adatron
KKT	Karush-Kuhn-Tucker
LBP	Local Binary Pattern

LCIS	Lobular Carcinoma In Situ
LDA	Linear Discriminant Axis
LoG	Laplacian of the Gaussian
LIBSVM	Library for Support Vector Machine
MATLAB	Matrix Laboratory
MCC	Microcalcification Cluster
MIAS	Mammographic Image Analysis Society
MLP	Multi-Layered Perceptron
MRI	Magnetic Resonance Imaging
MS-DOS	Microsoft Disk Operating System
MSE	Mean Square Error
NCR	National Cancer Registry
OS-ELM	Online-Sequential Extreme Learning Machine
PCA	Principle Component Axis
PPV	Positive Predictive Value
QP	Quadratic Programming
RBF	Radial Basis Function
RF	Random Forest
RFE	Recursive Feature Elimination
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
RLS	Recursive Least-Square
RMSE	Root Mean Square Error
SD	Standard Deviation
SGLDM	Spatial Gray Level Dependence Method
SLFN	Single Layer Feed-forward Neural Network
SMO	Sequential Minimal Optimization
SOM	Self-Organizing Map
SRG	Seeded Region Growing
SRM	Structural Risk Minimization
SSL	Sequentially Sorted List
STREL	Structuring Element
SVC	Support Vector Classification
SVM	Support Vector Machine
SV	Support Vector

TN	True Negative
TP	True Positive
TPF	True Positive Fraction
UCI	University of California Irvine
UMMC	University Malaya Medical Centre
US	Ultrasonography
USD	United States Dollar
VC-dimension	Vapnik—Chervonenkis dimension

CHAPTER 1

INTRODUCTION

The discovery of a lump in the breast is one of the most frightening and feared health problems women can face. This is due to the fact that breast cancer is the most common cancer to afflict women in most parts of the world (Boyle & Levin, 2008).

The aim of this research is the development of a reliable tool for the detection of breast cancer using digital mammography images. Image processing, data mining and machine learning techniques constitute the proposed framework of this thesis. The initial sections in this Chapter give an overview of breast cancer and the problems and challenges faced in breast cancer detection. In the later sections of this chapter the research motivation, objectives, contributions and an outline of the proposed framework in this thesis is presented. Finally, the structure of the rest of this thesis is illustrated in Section 1.6.

1.0 Overview

Breast cancer is the major cause of fatality among all cancers for women aged between 35 to 54 years (Verma & Zakos, 2001) and continues to be the leading cause of non-preventable cancer deaths (Kopans, 1989). Breast cancer is a serious problem in the United States, the incidence of which continues to rise (Bassett *et al*., 1997). A study made by the American Cancer Society (ACR) in 2003 estimated that in the United States between 1 in 8 and 1 in 12 women develop breast cancer during their lifetime (American Cancer Society, 2003a). According to these

statistics, on average, every 15 minutes five women are diagnosed with breast cancer, and one woman dies of this disease (Basett *et al.*, 1997).

According to statistics by the ACR, between 1973 and 1999, breast cancer incidence rates have increased by approximately 40 percent (American Cancer Society, 2003a). However, between 1989 to 1995 breast cancer mortality rates declined by 1.4 percent per year and by 3.2 percent afterwards. These declines have been attributed in large part, to early detection (American Cancer Society, 2003b). In the year 2009, the ACR estimated 192,370 new cases of invasive breast cancer amongst women, as well as 62,280 cases of in situ breast cancer. Moreover, the ACR estimated that in 2009, approximately 40,170 women were estimated to die from breast cancer (American Cancer Society, 2009). Figure 1.1 indicates the incidence and mortality statistics of the most prevalent cancers worldwide in 2008 amongst females (Boyle & Levin, 2008), of which breast cancer has one of the highest statistics as compared to other cancers amongst females.

Breast cancer is the major cause of death for women in Europe (Ferlay *et al.*, 2007). In Europe, breast cancer represents 19 percent of cancer deaths and 24 percent of all cancer cases (Esteve *et al.*, 1993). In 2006 in Europe the most common form of cancer diagnosed amongst females was breast cancer, with 429,900 cases (28.9 percent of all cancer cases) and 131,900 cancer deaths. In the European Union (EU25), breast cancer is the most common cancer with 319,900 cases (30.9 percent of all incident cases). In women, breast cancer is the major cause of mortality with 85,300 cases (Ferlay *et al.*, 2007).

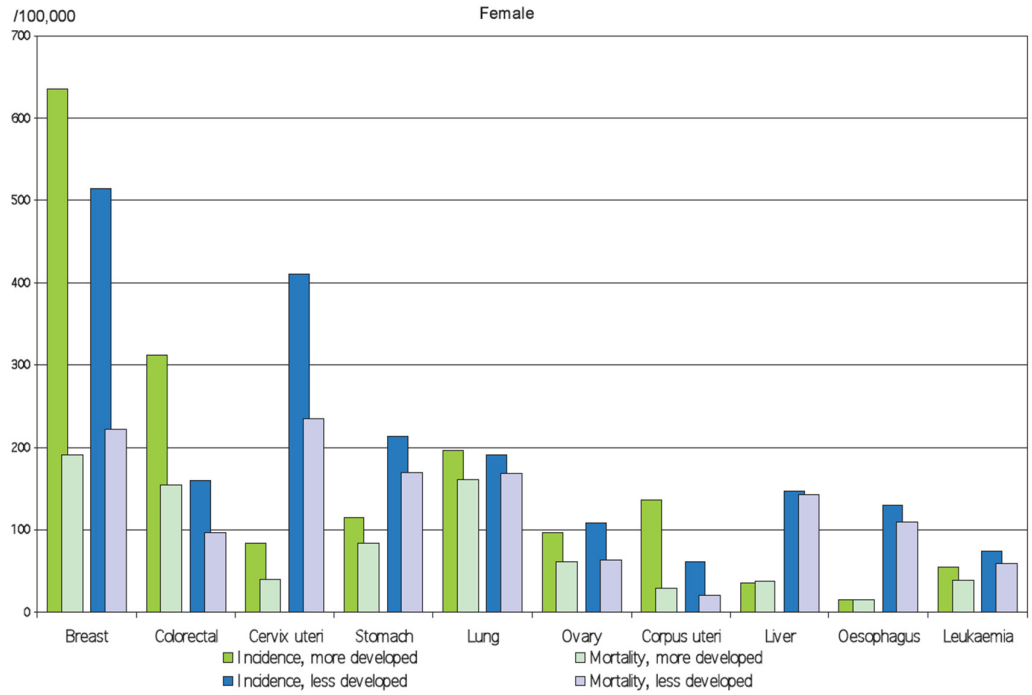


Figure 1.1: Incident and mortality statistics of most common cancers worldwide in 2008 amongst females (Boyle and Levin, 2008)

A study made on 100,000 women from 1995 to 1998 in the European Union, indicated that for breast cancer approximately 39 deaths per year occur regardless of the age against 40 deaths per year from 1985 to 1989, showing a change in rate equal to -2.1 percent. The favourable trend is due to the advancements, screenings, and the diagnosis of cancer at early stages (Levi *et al*. 2003). The chances of survival significantly grow if the illness is detected at an early stage. A reduction in breast cancer mortality for European countries in the 1990s was reported by several research groups such as Levi *et al.*, (2005) and Tyczynski *et al*. (2004). These declines in the mortality rate have been attributed to the combined effect of early detection of cancer and improving treatment. With the introduction of digital mammography screening programmes throughout Europe the reduction in breast cancer mortality (IARC Handbooks of Cancer Prevention, 2002) is expected to decrease.

In Malaysia, the most common type of cancer diagnosed is breast cancer (18 percent of all cancer cases). Breast cancer is the most common type of cancer among Malaysian women. In a 3 year period from 2003 to 2005, a total of 11,952 new cases were reported to the National Cancer Registry (NCR), Malaysia. Breast cancer accounted for 31.3 percent of the total number of new cases in women, with a similar percentage in each of the major ethnic groups; Malays (33.6 percent), Chinese (30.6 percent) and Indians (31.2 percent). The age-standardized rate (ASR) for females was 47 per 100,000 women (Lim *et al*., 2008). Figure 1.2 indicates the age specific cancer incidences per 100,000 population in Peninsular Malaysia from 2007 to 2008.

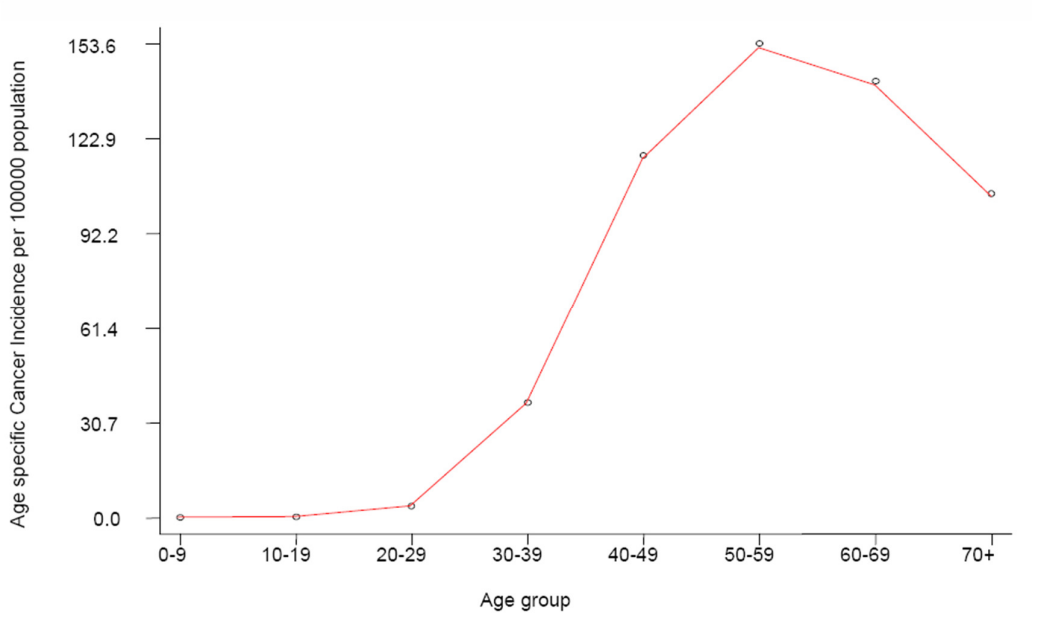


Figure 1.2: Age specific breast cancer incidences per 100,000 population in Peninsular Malaysia 2007 to 2008 (Lim *et al*., 2008)

The NCR in Malaysia in a report by Lim *et al*., 2008 estimated that the peak incidence of breast cancer occurred in women between the ages of 50 to 60 years, except in Indian women where the incidence peaked after the age of 60 years. The

incidence of breast cancer in Chinese women (ASR of 59.9 per 100,000 women) was higher compared to Malay women (ASR of 34.9) and Indian women (ASR of 54.2). Moreover, in the same report, it was noted that the age-standardized incidences amongst Malaysian women was lower as compared to several Western countries: USA (92.1), Canada (78.5), England (74.4), South Australia (80.8), Netherlands (85.6) and Denmark (81.3), but higher compared to some Asian countries such as: Beijing (24.6), Hiroshima (36.6), Chennai (23.9) and Seoul (20.8). Figure 1.3 provides international comparisons for age-standardized incidences of breast cancer per 100,000 women from 2007 to 2008.

Survival through breast cancer is stage-dependent and the best survival is observed when the cancer is diagnosed at an early stage. Mammography is currently the best technique for reliable detection of early, potentially curable breast cancer (Cardenosa, 1996), because it can detect cancerous cells such as: mass lesions, microcalcification clusters (MCCS) and other suspicious anomalies up to two years before they are palpable. Mammography has proven to be useful in detecting cancerous cells that may be unnoticeable by physical examination (Palmer *et al*. 2003).

As breast cancer incidents have increased during the recent decades, breast cancer mortality has gradually reduced for women of all ages (Sickles, 1997). This trend in mortality reduction is due to the adoption of mammography screening (Sickles, 1997), (Anttinen *et al*., 1993), (De Koning *et al*., 1995), (Hendee *et al*., 1999), (Tabar *et al*., 1985), (Thurfjell *et al*., 1994), which allows the detection of cancer at the early stages and the improvements made in the treatment of breast cancer (Buseman *et al*., 2003).

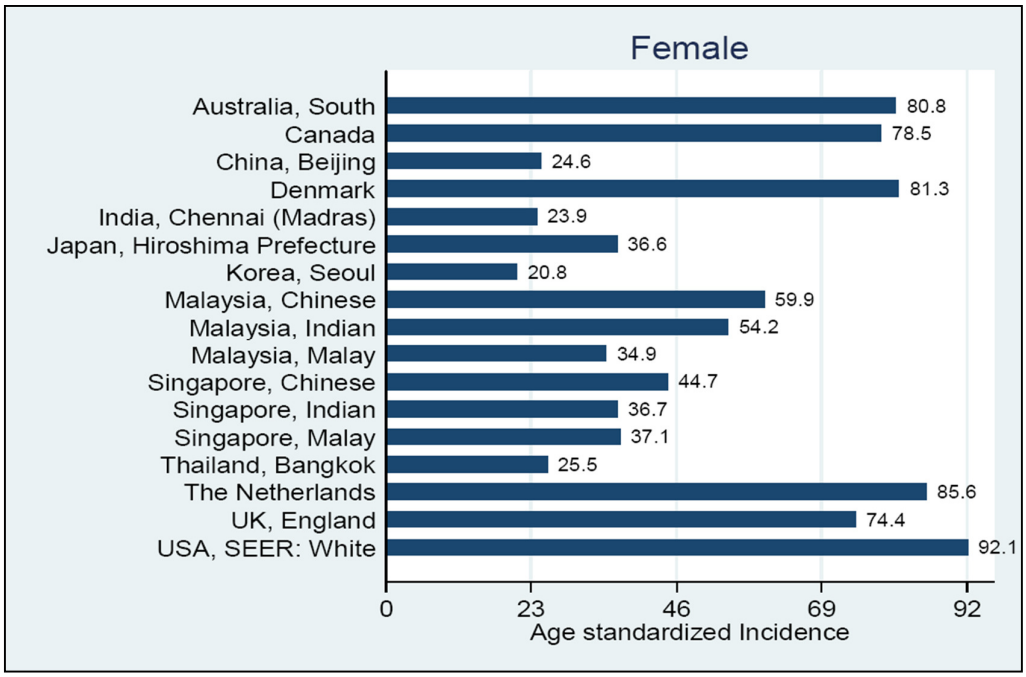


Figure 1.3: International comparisons — Age-standardized incidences of breast cancer per 100,000 population (Lim *et al.*, 2008)

Early stage breast cancers are associated with high survival rates. Thus, the key to surviving breast cancer is early detection and treatment. According to the ACS, when breast cancer is confined, the five-year survival rate is almost 100 percent. Breast cancer screening has shown to reduce breast cancer mortality. Currently, 63 percent of breast cancers are diagnosed at a localized stage, for which the five-year survival rate is 97 percent (American Cancer Society, 2003b). The high survival rates are attributed to the proper utilization of mammography screening as well as high levels of awareness of the disease symptoms in the population.

1.1 Problems and Motivation

Digital mammography is a relatively new technique for the early detection of breast cancer. It is based on accumulated density of tissues, that is, to detect shadows. This is the reason as to why digital mammography has been considered as an efficient tool for the detection of masses and MCCs (Khuzi *et al.*, 2009),

(Verma & Zakos, 2001), (Jiang *et al.*, 1999). The potential advantages of digital mammography are:

1. *Image Processing Capability*

   Clinical images presented optimally in digital format. (Appropriate contrast, density and edge enhancement can be achieved by separating the functions of image display and image recording).

2. *Possibility of Computer-Aided Detection Methods*

   Several tools for computer-aided detection and computer-aided diagnosis are under development. In the near future, these tools will provide cost-effective mammography screening, where double reading for radiologists' will be recommended.

3. *Picture Archiving and Management*

   The currently technology based on a storage phosphor system, is a sub-optimal technique because of low spatial resolution. In the near future, high quality digital mammography using full field digital technology will be available.

The major problem associated with mammogram screening programmes is the large percentage of missed cancers. Studies show that during screening, radiologists fail to detect 15 percent of breast cancers that are visible in retrospective studies (Goergen *et al.* 1997), (Bird *et al.* 1992). Moreover, when minimal signs are taken into account, estimates of missed cases increase to 50

percent (Timp *et al.*, 2004), due to errors of perception. Eye tracker studies have classified these mammogram screening errors into three main categories:

1. *Searching Error:* In this case the radiologist overlooks the abnormality. Eye tracker experiments show that foveal sight never reached the lesion.

2. *Detecting Error:* The lesion has been seen by the radiologist, but the visual dwell time was shorter than a certain threshold, for instance one second.

3. *Interpretation Error:* These lesions are consciously evaluated by the radiologist, but acted on with an inappropriate decision.

Without considering recorded eye movements, search and detection errors are those that occur when radiologists do not report the presence of a visible cancer and interpretation errors as those that occur when the tumor is reported but not considered actionable.

In order to identify if a breast tissue is cancerous, a biopsy is usually performed. During a biopsy the suspicious breast tissue is removed from the patient for a diagnostic examination. The breast tissue is removed by a surgical excision and is diagnosed by the pathologist. An abnormality, once detected in a breast tissue, can be classified as either *benign* (not cancerous) or *malignant* (cancerous) (Hadjiiski et al., 1999), (Wei *et al*., 2005).

Although digital mammography has proven to be an efficient tool for detecting breast cancer, the interpretation of mammograms however requires the skills and

experience of expert radiologists (Khuzi *et al*., 2009). Clinical studies have shown that the Positive Predictive Value (PPV) (ratio of the total breast cancers found to the total number of biopsies) is only 15 to 30 percent (Kopans, 1992), (Adler & Helvie, 1992), (Moskowitz, 1989), (Giger and MacMahon, 1996). The performance of a medical diagnostic test is typically measured using the Receiver Operating Characteristic (ROC) curve analysis as presented in Section 3.3.6, where the four performance measures: true positive (TP), false positive (FP), true negative (TN) and false negative (FN) measure the *sensitivity* and *specificity* of the tested samples with malignant samples being the $positive$ $(+ve)$ class and benign samples being the $negative$ $(-ve)$ class. Poor mammographic image quality, physician eye fatigue, subtle nature of radiographic findings and other sources may cause incorrect identification (misclassification) of a malignant abnormality as benign, generally referred to as a false positive (FP) (Kocur *et al*., 1996), (Fogel *et al*., 1998). A diagnostic model with this problem results in a bias when implemented in the diagnosis of breast cancer. Thus, the misclassification of a benign/malignant patient is defined as Type I/II error respectively (Fisher, 1936).

Mass lesions (or masses) and microcalcification clusters (MCCs) are the two most important radiographic indications related to breast cancer, as they are present in 30 to 50 percent of all cancers found mammographically (Sickles 1984), (Sickles, 1986). The detection of mass lesions is a challenging task, because:

- Lesions are normally hidden or found in the dense glandular area of the breast tissue, which is sometimes difficult to distinguish due to the variation in shape, size and dimension (Wei *et al*. 2005).

- Lesions are usually indistinguishable from the surrounding tissue because their features (heuristics) can be obscured as they are similar to the normal inhomogeneous breast tissues (Bozek *et al*., 2008).

Similarly the detection of MCCs is also challenging, because:

- MCCs are calcium deposits of very small dimension and appear as a group of granular bright spots in a mammogram (Wei *et al*., 2005). They appear as tiny circular objects, which can be described as irregular, granular or linear and can vary in size form 0.1mm to 1mm having an average diameter about 0.3mm (Diyana *et al*., 2002). Small MCCs ranging from 0.1 to 0.2mm can hardly be seen on the mammogram due to their superimposition on the breast parenchymal texture and noise (Diyana *et al*., 2003a).
- MCCs often appear in an inhomogeneous background describing the structure of a breast tissue. Some parts of the background have features such that the dense tissues may be brighter than the MCCs in the fatty tissue (Diyana *et al*., 2003b). This is due to a large amount of absorbent tissue (mainly fibroglandular tissue) in the dense breast image, so the image contrast needs to be decreased.
- Some MCCs have low contrast compared to their background such as breast tissues, blood vessels, mammary glands and fat. In other words, the intensity and size of the MCCs can be very close to noise or an inhomogeneous background (Diyana *et al*., 2002).

Thus, the masses and MCCs are relatively difficult to detect and can be overlooked by radiologists in mammography screening. Considering the traumatic nature, cost

of biopsy and the relatively difficult task for radiologists to interpret mammograms, it is desirable to develop computer-based methods which can accurately distinguish between benign and malignant abnormalities (Mudigonda *et al.*, 2000). In addition, it is important to increase the positive predictive value (PPV) without reducing the sensitivity of breast cancer detection. The use of double reading by two or more radiologists has shown to improve the sensitivity, but it also increased the cost of the mammogram screening process (Khuzi *et al*., 2009), (Mousa *et al.*, 2005).

In order to improve the biopsy yield ratio, computer-aided methods are desirable for the detection of masses and MCCs and for the further classification of the detected abnormality as benign or malignant (Mudigonda *et al.*, 2001). Such methods provide ease in performing initial mammogram screening and second reading of mammograms to help radiologists in analyzing difficult cases, especially in deciding on biopsy and follow-up recommendations (Mudigonda *et al.*, 2000). Using a computer-aided detection scheme, radiologists can incorporate the output from the computer into their decision. Several recent studies have shown that computer-aided detection improves the radiologists' ability in differentiating between benign and malignant abnormalities (Giger, 1999), (Jiang *et al*., 1996), (Jiang *et al*, 1999), (Wu *et al.*, 1993), (Huo *et al.*, 2000), (D'Orsi *et al.*, 1992), (Baker *et al.*, 1996), (Chan *et al.*, 1999).

Computer-aided detection methods are a combination of image processing and machine learning techniques (Hutt, 1996). Mammogram processing typically uses image processing techniques for the purpose of suppressing artifacts and labels, eliminating digitization noises and enhancing mammograms for optimal viewing,

mainly for detection of mass lesions and MCCs (Mutihac *et al.,* 1998), (Strickland *et al.,* 1996), (Cernadas *et al*., 1996), as presented in Section 3.3 of this thesis. Many studies provide evidence that radiologists perform better on computer enhanced images (Aylward *et al*., 1998) (Netsch *et al*., 1998) to transform mammograms in such a way that they can be printed or examined on a monitor optimally (Karssemeijer & te Brake, 1996), (Bynd *et al*., 1997).

Machine learning techniques are typically applied in computer-aided detection schemes for the purpose of pattern classification (Hutt, 1996). During classification, features estimated from the Region of Interest (ROI) are used for the training and testing (validating) a learning machine, as presented in Chapter 4. A classifier trained on known abnormalities (mass lesions and MCCs) combines the selected features and uses confidence measures to indicate if the tested sample is malignant or benign. Several classification techniques have been investigated for the detection of benign and malignant abnormalities in mammograms during the last decade, as reviewed in Section 3.2. These classification techniques include: Support vector machines (SVMs) (Wei *et al*., 2005), Artificial Neural Networks (ANNs) (Wu *et al*., 1992), (Chan *et al*., 1995a), (Jiang *et al*., 1996), (Sahiner *et al*., 1996), (Chan *et al*., 1997), (Huo *et al*., 1998), (Papadopoulos *et al*., 2002), (Zhang *et al*., 2005), Linear Discriminant Analysis (LDA) (Chan *et al*., 1995b), (Zhang *et al*., 2005), Convolutional Neural Networks (CNNs) (Sahiner *et al*., 1996) and the k-Nearest neighbor (Veldkamp *et al*., 2000). Other machine learning techniques include the use of statistical-based models and the general framework of Bayesian image analysis, which was developed by Karssemeijer (1993).

1.2 Research Objectives and Scope

The aim of this research is to work out on the second potential advantage of digital mammography discussed in Section 1.1, that is, computer-aided detection. This research focuses on developing a framework of algorithms using image processing and machine learning techniques for the detection of malignant and benign abnormalities in digital mammography images. Prior to that, related techniques for image processing and machine learning will be reviewed in order to identify the most suitable approach for the detection of malignant and benign abnormalities, which includes the detection of mass lesions and MCCs.

The goal of this research is to increase the diagnostic accuracy of image processing and machine learning techniques for optimum classification between malignant and benign abnormalities as well as to the reproducibility of mammographic interpretation. In order to achieve the goal of this research, the following research objectives are set:

1. To investigate and apply existing image processing algorithms and machine learning techniques in order to detect mass lesions and MCCs in digital mammograms.

2. To develop a system using the investigated techniques, for the classification of malignant and benign abnormalities, using a combination of image processing algorithms and machine learning techniques.

3. To apply the identified algorithms and techniques in order to reduce the number of misclassified malignant cancers, namely, false positives (FPs) (see Section 3.3.6).

4. To verify the reliability and accuracy of the developed system using the datasets from different sources.

5. To perform a comparative study for identifying the most suitable machine learning technique for the classification of benign and malignant abnormalities.

The proposed framework in this research relies mainly on image processing algorithms and machine learning techniques. For a deeper understanding of the proposed framework refer to Section 1.4.

1.3 Research Significance and Contribution

In addition to the problems discussed in Section 1.1 for the detection of masses and MCCs, in digital mammograms, it is observed that most mass lesions have a stellate appearance in mammograms. Moreover, their central masses are typically irregular with ill-defined borders and can vary in size from a few millimetres to several centimetres in diameter. Due to these problems and the problems presented in Section 1.1, the detection of malignant and benign masses and the detection of MCCs has become a challenging task.

Applying a combination of image processing and machine learning techniques for the detection of abnormalities in digital mammograms requires feature

calculation/estimation from the Region of Interest (ROI), namely, the abnormal region. In general, it is difficult to determine the size of the neighbourhood (pixels) or the ROI that should be used to calculate the relevant features from the abnormal regions (masses and/or MCCs). If the size of the ROI is too large, small masses and/or MCCs may be missed, while if the size of the ROI is too small, parts of large masses and/or MCCs may be missed. This poses a challenging task in the detection of breast cancer. Thus, the primary contribution of this research is:

- To determine the most suitable ROI (neighbourhood) size of mass lesions and MCCs for the purpose of feature computation (extraction).

This specifically addresses the difficulty of predeterming the ROI size for the purpose of feature estimation (extraction). The detection of masses and/or MCCs is regarded as one of the hardest to solve in the field of recognition of objects into images (see Section 1.1). The difficulty in carrying out research, such as this, lies not only in the process itself, because even radiologists find it challenging to identify masses and MCCs given their variability in shape, size and dimension. Thus, the secondary contribution of this research is:

- To demonstrate that advanced machine learning techniques, namely, Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) can effectively solve pattern classification problems.

The secondary contribution of this research provides the basis for conducting a comparative analysis between different machine learning technologies, such as ANNs and SVMs which complies with the fifth research objective in Section 1.2.

1.4 Research Methodology and Proposed Approach

The literature discussed in Chapters 2, 3 and 4 of this thesis presents the state-of-the-art image processing and machine learning algorithms applied for the detection of malignant and benign abnormalities in digital mammograms. Figure 1.4 illustrates an overview of the basic framework developed in this research. For a detailed illustration of the proposed framework and applied techniques, refer to Figure 3.1 and Figure 5.1. The inputs into the proposed framework are digital mammogram images, whereas the output of the system indicates that the abnormality in the ROI of the input image is malignant or benign.

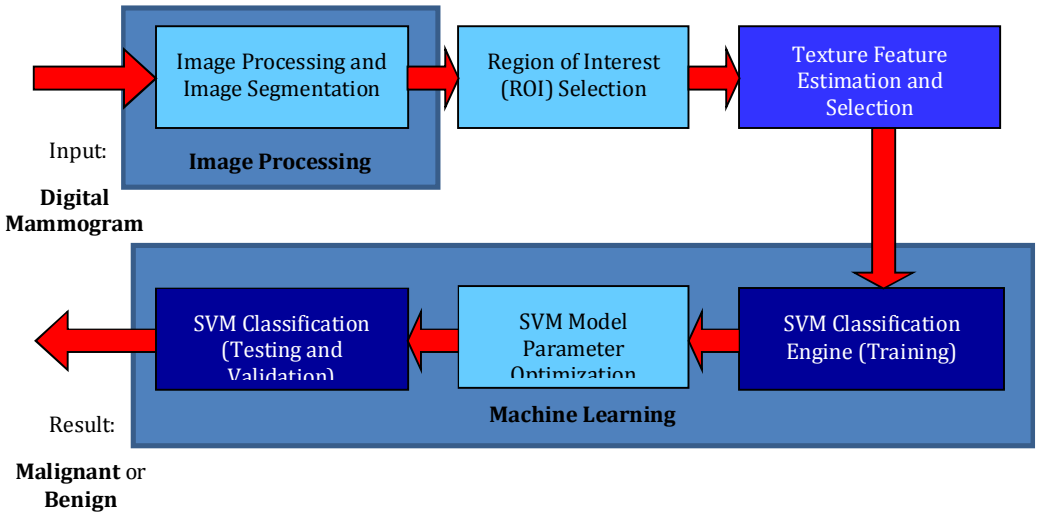


Figure 1.4: Overview of the proposed framework for classification of malignant and benign abnormalities in digital mammograms

The literature reviewed in Chapter 2 and Section 3.3 indicates that digital image processing techniques need to be applied to the mammogram images for the purpose of noise removal and pectoral muscle segmentation. Usually, mammogram preprocessing techniques includes: noise removal, background suppression and artifact/wedge and label removal as discussed in Section 3.3.1. Since the breast profile needs to be optimally segmented from the background region, the pectoral

muscle is also suppressed from the mammograms since it may bias the procedures in the detection of malignant and benign abnormalities as discussed in Section 3.3.2. The ROIs (benign and malignant samples with mass lesions and/or MCCs) are extracted from the segmented mammogram images using the *Ground Truth* (GT) markings from the radiologists' interpretation.

Digital mammograms are known to contain prominent textural information regarding the shapes and sizes of masses and MCCs. Texture based features are known to increase the performance of machine learning algorithms such as ANNs and SVMs (Belotti *et al*., 2006), (Makinacı *et al*., 2005), (Jirari *et al*., 2005). The approach proposed in this research uses texture features for the purpose of pattern classification, as discussed in Section 3.3.3. Thus, texture descriptors are computed from the ROIs representing the malignant and benign samples.

As discussed in Section 1.3, SVMs can effectively solve binary classification problems using noisy data. The theoretical concepts of machine learning and SVMs are discussed in Chapter 4. Binary classification results obtained from the SVM testing in Figure 1.4 are analyzed using Receiver Operating Characteristic (ROC) curves presented in Section 3.3.6, to estimate the classification accuracy of the SVM for unseen samples.

1.5 Benefits of Image Processing and Machine Learning Techniques

Digital mammography leads itself well to breast cancer detection, where image processing algorithms enable computers to indicate suspicious areas of the breast that contain masses, MCCs or other prominent signs of breast cancer.

Image processing algorithms have the potential to increase the diagnostic accuracy by reducing the number of FPs (misclassified malignant cancers), through the application of image segmentation. Using segmentation, the unimportant parts in the ROIs (the breast region near to the breast border and the boundary between the segmented pectoral muscle and the breast tissue) can be eliminated, thus making feature estimation and calculation more accurate, as discussed in Section 5.4.1.1. The benefits obtained of using the proposed framework in Figure 1.4 are as follows:

1. The proposed system will aid radiologists in their diagnosis by indicating suspicious abnormalities in mammograms. Thus, the system will act as a second reader after the radiologists.

2. The proposed system will substantially reduce the number of false positives (FPs), (see Section 3.3.6), which will eliminate the need of performing unnecessary biopsies and save cost.

3. This system will reduce patient examination time by inspecting mammograms and reporting the findings within a few seconds.

In breast cancer diagnosis, the weakest link has always been the radiologists, since it is the radiologists who must find mass lesions and MCCs in order to make a diagnosis. So, the radiologists can refer to this system for a second opinion as it is difficult to distinguish between malignant (cancerous) from benign (non-cancerous) tissues due to their similar nature and visual features (Veldkamp *et al*., 2000), (Rahbar *et al*., 1999), (Jiang *et al*., 1998).

1.6 Thesis Overview

This thesis is arranged in a methodical manner. It is organized into seven chapters comprising of this introduction chapter and six further chapters as follows.

Chapter 2 discusses breast cancer detection and digital mammography. General information about the structure and functions of the breast, breast tumors and literature regarding breast cancer screening is presented at first. Next, background issues concerning different imaging modalities such as digital mammography, ultrasonography (US) and magnetic resonance imaging (MRI) are reviewed and discussed. Towards the end of this chapter, the analysis and interpretation of digital mammograms using the BI-RADS lexicon is discussed.

Chapter 3 presents and discusses the background literature on the computer processing of digital mammograms. In Section 3.1, computer-aided detection systems are introduced with the literature review of computerized breast cancer detection techniques presented in Section 3.2. Section 3.3 highlights and identifies the key techniques and algorithms used in this research to develop a framework for the computerized detection of breast cancer. Section 3.4 presents the fundamental concepts of digital image processing with emphasis on image segmentation techniques used in digital mammography applications. Lastly, Section 3.5 emphasizes on the use of texture-based analysis for the purpose of feature extraction in pattern classification problems.

In Chapter 4, an overview of pattern recognition is presented, with particular emphasis on a specific machine learning technique, namely, the Support Vector Machine (SVM). SVMs will be used intensively in this research. The reason for

using SVM as the main machine learning technique for this research is discussed in Sections 1.3 and 3.3.5. Section 4.1 presents some introductory notions regarding the theoretical concepts of learning machines. Section 4.2 introduces the fundamental concepts of the statistical learning theory and presents the mathematical formulation of the SVM developed by Vapnik (1998) which describes the statistical aspects of automated machine learning. Towards the end of this chapter, Section 4.3 presents the theoretical concepts of ANNs whereas Section 4.4 discusses a Recursive Feature Elimination (RFE) technique used for the selection of the optimal subset of texture features for the learning machine (SVM).

Chapter 5 presents the modeling of the framework (system) proposed in Chapter 1 for the classification of benign and malignant abnormalities in digital mammograms. As discussed in Chapter 3, the proposed framework is composed of two main techniques, namely, image processing and machine learning. The image process and machine learning techniques identified in Section 3.3, are applied in the proposed system in this chapter, which are discussed in Section 5.1 and 5.2. The modeling of the system consists of three main stages, namely: Mammogram Image Processing (Section 5.3), Texture Feature Extraction and Selection (Section 5.4) and Classification Engine (Section 5.5). Sections 5.3 through 5.5 describe each stage in detail during the development of the proposed framework.

Chapter 6 presents the experimental results of the developed system in Chapter 5. Section 6.1 presents and discusses the SVM training results relative to the memorization and learning of the binary SVM classifier. Section 6.1 also presents and discusses the SVM testing and validation results for unseen samples. In order to perform a comparative research, Section 6.2 presents the experimental results

obtained after evaluating the developed framework using different machine learning algorithms other than the SVM. The experimental results of the compared machine learning models are discussed in the last part of Chapter 6.

Chapter 7 concludes and summarizes the research contributions made. The achievements and objectives of the research with respect to the experimental results obtained are highlighted along with the key findings and significance of the research. This chapter also discusses the impact and significance of the developed system to radiologists and hospitals for mammography screening and interpretation. Radiologists and clinicians will benefit from the developed system as it will assist them in their diagnosis by acting as second readers.

CHAPTER 2

DIGITAL MAMMOGRAPHY

2.0 Overview

This chapter discusses breast cancer detection and digital mammography. General information about the structure and functions of the breast, breast tumors and literature regarding breast cancer screening is presented at first. Next, background issues concerning different imaging modalities such as digital mammography, ultrasonography (US) and magnetic resonance imaging (MRI) are reviewed and discussed. Towards the end of this chapter, the analysis and interpretation of digital mammograms using the BI-RADS lexicon is discussed.

2.1 Breast Anatomy and Cancer

The most important anatomical structures of the breast are shown in Figure 2.1. The breast consists of two components. The first component is concerned with milk production and is known as the epithelial component. The second component consists of fat and connective tissue, which supports and protects the structure of the breast (Bassett *et al.*, 1997).

The epithelial component of the breast consists of a tree-like branching pattern of milk ducts that come together at the nipple. The leaves of this tree are formed by the lobules which are the secretory units of the breast. Each lobule consists of a number of acini connecting to an intra-lobular duct. The acini are composed of two types of cells, namely, the epithelial and myo-epithelial. The epithelial cells secrete

a variety of glyco-proteins and during lactation they also produce milk. The myo-epithelial cells are capable of contracting during breastfeeding. Each intra-lobular duct connects with an extra-lobular duct, and this together with the lobule, is called the terminal ductal lobular unit (Chu *et al.*, 1988).

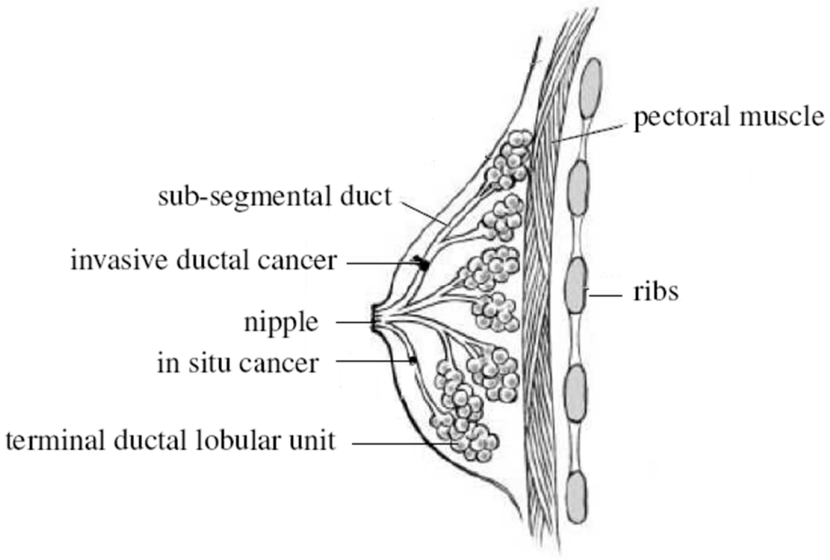


Figure 2.1: Anatomy and structure of the female breast

The extra-lobular ducts of the breast link together and form sub-segmental ducts, which in turn form the segmental ducts. These ducts drain milk from different segments or lobes of the breast. In total, the breast consists of 15 to 20 lobes, which are roughly pyramidal in shape with the apex directed towards the nipple. The non-epithelial component of the breast consists mainly of fatty tissue. There are no muscles in the actual breast, but there are a series of muscles behind and underneath the breasts. These muscles work together with a ligament called Cooper ligament to support the weight of the breasts (Chu *et al.*, 1988).

Breasts contain lymph vessels, which are very important in fighting against diseases in the body. Lymph is a clear fluid that contains tissue fluid, waste

products, as well as immune system cells. The lymph system consists of lymph nodes and lymph vessels that transport the lymph to the lymph nodes. Most of the lymph vessels that go through the breast carry the lymph to the lymph node underneath the arm pit, called axillary nodes. The other lymph vessels carry the lymph to the lymph nodes which are inside the chest, called the internal mammary nodes, or the lymph nodes above or below the collarbone, called the supraclavicular or infraclavicular nodes respectively. Lymph veins can also carry possible diseases to lymph nodes which might increase the spread of the disease, for example breast cancer (malignant) cells (Boyle, & Levin, 2008).

Breast cancer is the major cause of fatality among all cancers for women aged between 35 to 54 years (Verma and Zakos, 2001) and continues to be the leading cause of non-preventable cancer deaths amongst females, as indicated in Figure 1.1. Breast cancer is developed when the cells of the breast become abnormal (malignant) and spread without order or control. The malignant cells then form a tissue and turn into a tumor. This tumor typically grows into nearby tissues or breaks away and enters the bloodstream or lymphatic system which can affect other organs. The spreading of breast cancer is generally referred to as metastasis (Kopans, 1989).

The most common and effective method for detecting breast tumors in their early stages is by performing mammogram screening. Mammography is currently the most effective modality used to detect tumors in the breast tissue (Sickles, 1997), (Anttinen *et al*., 1993), (De Koning *et al*., 1995), (Hendee *et al*., 1999), (Tabar *et al*., 1985), (Thurfjell *et al*., 1994) that can indicate potential clinical problems, such as the: asymmetries between breasts, architectural distortion, confluent densities

associated with benign fibrosis, microcalcification clusters (MCCs) and mass lesions. By far, the two most common features that are typically associated with breast tumors are MCCs and mass lesions, which are discussed in the following sections.

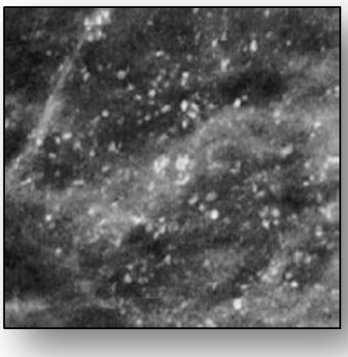


Figure 2.2: Microcalcification clusters (MCCs) in a breast tissue

2.1.1 Calcifications

Calcifications are small mineral (calcium) deposits within the breast tissue that appear as localized high intensity regions in the mammogram. There are two types of calcifications: microcalcifications and macrocalcifications. Macrocalcifications are coarse, scattered calcium deposits. These deposits are usually associated with benign conditions and rarely require a breast biopsy. Microcalcifications on the other hand are isolated calcium deposits that normally appear in clusters or are found embedded in a lesion. Individual microcalcifications typically range in size from 0.1 to 1.0mm with an average diameter of about 0.5mm. A microcalcification cluster (MCC) is typically defined to be at least three microcalcifications within a $1cm^2$ region, as shown in Figure 2.2. About 30 to 50 percent of non-palpable cancers are initially detected due to the presence of MCCs (Feig & Yaffe, 1995). Similarly, in a large majority of the ductal carcinoma in situ (DCIS) cancers, MCCs are present (Monsees, 1995).

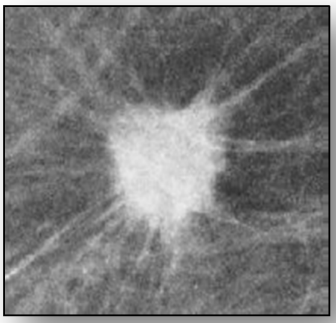


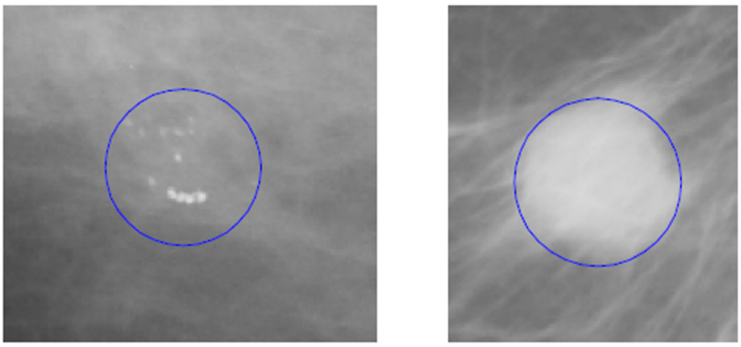
Figure 2.3: Mass lesion in a breast tissue

2.1.2 Mass Lesions

Breast tumor is often represented as a mass lesion with or without the presence of MCCs (American Cancer Society, 2003a). A cyst, which is a non-cancerous collection of fluid, may appear as a mass in the film. However, ultrasound or fine needle aspirations can distinguish the difference. The similarity in intensities with the normal tissue and morphology with other normal textures in the breast makes it more difficult to detect masses compared with calcifications (Feig & Yaffe, 1995). The location, size, shape, density, and margins of lesions are useful for the radiologist in evaluating the likelihood of a cancer (Evans, 1995). Most benign masses are well circumscribed, compact, and roughly circular or elliptical, as shown in Figure 2.3. Malignant lesions usually have a blurred boundary, irregular appearance and sometimes are surrounded by a radiating pattern of linear spicules (Evans, 1995). However, some benign lesions may have a spiculated appearance or blurred periphery.

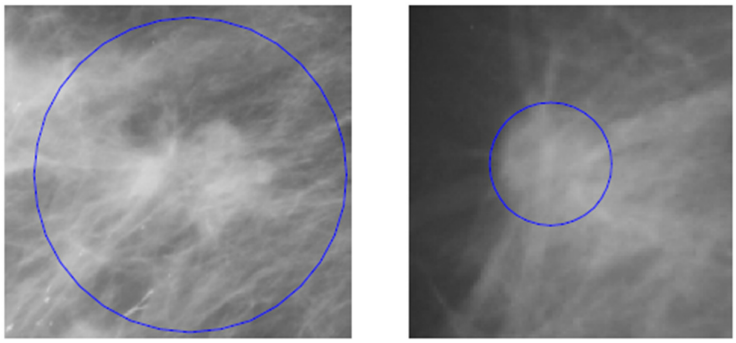
Mass lesions and MCCs are abnormal regions in mammograms. Examples of mammograms with MCCS and masses are shown in Figure 2.4 and Figure 2.5, with the *Ground Truth* (GT) data (in Section 5.2.1) superimposed. The GT data for a

mammography dataset are the radiologist's findings in the diagnosis of mammographic abnormalities which are the location, size and shapes of suspicious masses and/or MCCs found, as shown in Figure 5.5. As observed in Figures 2.4 and 2.5, there are several different lesion types and lesions can either be malignant or benign. Malignant tissues indicate cancer, whereas benign tissues indicate non-cancerous cells, i.e. abnormal tumors. The following section discusses cancerous and non-cancerous breast tumors in detail.



(a) (b)

Figure 2.4: Examples of abnormal regions of mammograms
(a) Microcalcifications (b) Circumscribed mass



(a) (b)

Figure 2.5: Examples of abnormal regions of mammograms
(a) Spiculated mass (b) A mass lesion classified as miscellaneous

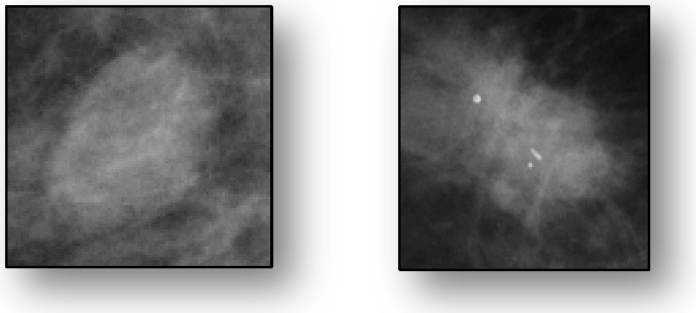
2.2 Breast Tumors

Breast tumor is identified by different names, depending on where it starts in the woman's breast. Scientists are not sure of the exact cause of breast cancer, but they have identified high risk factors for this disease. The most common factors include: age, family history and personal history. The most common symptom is a painless lump in the breast. At times, a painful lump in the breast turns out to be cancer (malignant). One or more lumps in the armpit can be a symptom of breast cancer; however, they can also be due to non-cancerous (benign) conditions. The bleeding from the nipple can indicate the presence of cancer, especially if the bleeding occurs from one breast only (Buseman *et al*., 2003). A more difficult symptom to be identified is the thickening of the tissue in the breast. Any changes in the breast size or shape can be due to cancerous (malignant) or non-cancerous (benign) conditions. Although benign conditions in the breast can be understood by their symptoms, biopsies need to be performed to understand if the irregularity in the breast is non-cancerous or otherwise. Other symptoms indicating the possibility of breast cancer are: redness of the skin over a portion of the breast, redness or scaliness of the nipple or any nipple pain or retraction (nipple turning inward), an orange peel appearing on the skin and dimpling of the skin (Basett *et al.*, 1997).

In our research, we distinguish-two types of breast abnormalities, namely, *benign* and *malignant*. The following sections discuss cancerous and non-cancerous breast tumors in detail.

2.2.1 Non-Cancerous Breast Tumors

Non-cancerous (benign) tumors of the breast comprise fibro-adenoma, duct papilloma, adenoma and connective tissue tumors. The most common benign breast tumor is the fibro-adenoma. This tumor is a combined product of both connective tissue and epithelial cells in the breast. Most benign masses are circumscribed due to the absence of infiltration (Palmer *et al*. 2003).



(a) (b)

Figure 2.6: Appearance of breast lesions
(a) Characteristic example of a benign mass lesion
(b) A benign mass which presents as a malignant lesion

Figure 2.6(a) shows a characteristic example of a benign mass lesion. The shape is oval and the border is sharply delineated. Benign mass lesions may also be suspicious (presented as malignant lesions), as shown in Figure 2.6(b). Based on these reasons, mammographically it is difficult to distinguish between benign and malignant lesions (Khuzi *et al.*, 2009).

2.2.2 Cancerous Breast Tumors

Cancerous (malignant) breast tumors can be classified into two main types: (i) Non-Invasive breast cancer, and (ii) Invasive breast cancer. The following sections discuss cancerous breast tumors in detail.

2.2.2.1 Non-Invasive Breast Cancer

Non-Invasive—in situ—cancer consists of malignant cells that replace the normal epithelial cells, lining the ducts or lobules in the breast tissue. These malignant cells are confined to the basement membrane and have not yet invaded the breast stroma or lymphatics. The two forms of non-invasive breast cancer are: (i) *Ductal Carcinoma In Situ* (DCIS) and (ii) *Lobular Carcinoma In Situ* (LCIS) (Lu and Bottema, 2001), which are described as follows:

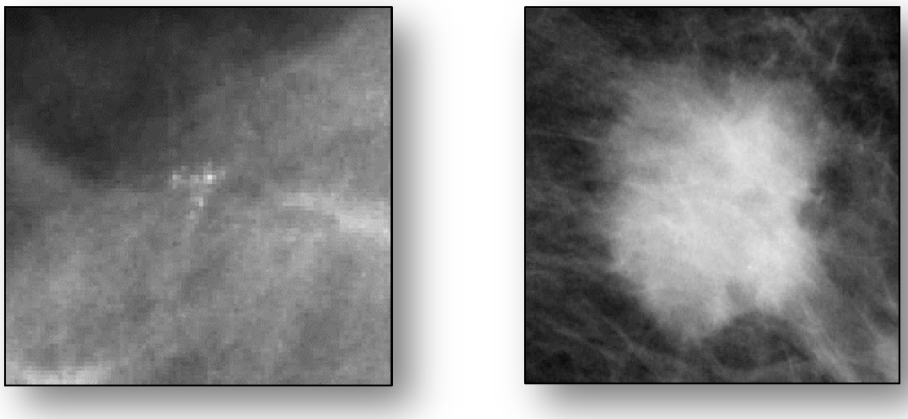
- **Ductal Carcinoma In Situ (DCIS)**: DCIS is a malignancy of the epithelial cells lining the lactiferous ducts (usually the terminal ducts) without penetration of the ductal basement membrane. The prognosis of untreated DCIS is not precisely known, as most patients are treated with mastectomy. It is estimated that about one-third to half of the untreated patients eventually will develop invasive cancer, usually in the same quadrant of the breast where the first lesion develops. Mammographically DCIS is often characterized by the presence of microcalcifications. When there is extensive fibrosis, DCIS may also present as a palpable mass (Lu and Bottema, 2001). Figure 2.7(a) shows an example of DCIS.

- **Lobular Carcinoma In Situ (LCIS):** In LCIS the lobules are expanded by a uniform population of small yet atypical cells. Usually this process obliterates the lumen of the acini. These atypical cells do not penetrate through the walls of the lobules. LCIS rarely gives rise to mammographic abnormalities. It is often found in biopsies that have been done for other reasons such as removal of benign lesions. LCIS is a risk factor for

developing breast cancer. The majority of patients are therefore managed by careful follow ups (Lu and Bottema, 2001).

2.2.2.2 Invasive Breast Cancer

Invasive breast cancer, also known as *infiltrating cancer*, occurs when malignant cells have spread beyond the ducts or lobules to other parts of the breast or body. Invasive cancers vary in size from less than 10mm in diameter to over 80mm, but are usually 20 to 30mm at presentation (Vitak, 1998).



(a) (b)

Figure 2.7: Appearance of breast lesions
(a) An example of ductal carcinoma in situ (DCIS)
(b) An infiltrative ductal malignant cancer with characteristic ill-defined and spiculated borders (invasive breast cancer)

Ductal carcinoma accounts for about 80 percent of all invasive breast cancer cases. These tumors are believed to arise from epithelial cells of the terminal ductal lobular unit. It is generally thought that ductal carcinoma starts as a DCIS. Less common types of breast cancer include: lobular carcinoma, medullary carcinoma,

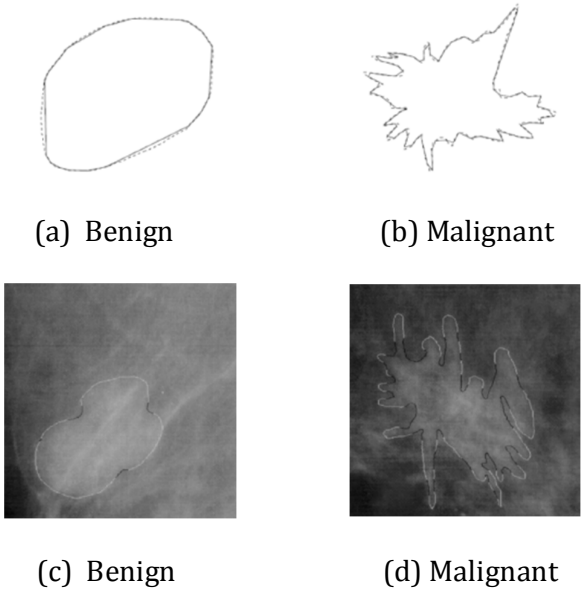
tubular carcinoma, mucinous carcinoma, cribriform carcinoma and papillary carcinoma (Popli, 2001). Figure 2.7(b) shows an example of an infiltrative cancer.

Breast cancers can infiltrate locally to the skin and the muscle, or metastasise to more distant sites via lymphatics or the bloodstream. The most common spread via lymphatics is to the axillary lymph nodes. Metastasis via the blood stream most frequently involves the lung and the liver, but adrenals and brains are also common sites for metastasis. When a woman has invasive breast cancer the prognosis depends among others on the histological grade and behavioral characteristics of the tumor and the presence of the metastatic spread (Vitak, 1998).

Considering histology, tumors can be graded based on their degree of differentiation. Well differentiated tumors often have a better prognosis than poorly differentiated tumors. Behavioral characteristics that influence prognosis include the growth rate and the receptor status of a tumor. Tumors with lower cell growth rates generally behave better. The presence of oestrogen receptors indicates that the tumor cells have a higher degree of functional differentiation resulting in a better prognosis (Popli, 2001). Tumor spread is also associated with a worse prognosis than when there is no evidence of metastasis. Although these factors may predict how individual cancers will behave, this has not led to an improvement of patient survival. Mammography screening on the other hand has proven to improve survival rates (Vitak, 1998), as discussed in Section 2.4.

2.3 Differentiating Between Breast Tumors

Although there are many types of breast abnormalities, it is possible to have a general differentiation between benign and malignant breast tumors using their boundary shapes with the surrounding breast tissue. This differentiation can be performed by examining spiculations on the malignant tumor, which can easily be observed using mammography or ultrasound techniques. Spiculation is a stellate distortion caused by the intrusion of breast cancer into the surrounding tissue and its existence is very important for tagging the tumor as malignant. There are many successful techniques based on mammography or ultrasound to quantify the degree of spiculations for a successful decision of differentiation between the benign and malignant type of breast tumors (Huang *et al.*, 2004). Figures 2.8(a) and 2.8(b) indicate a benign and malignant tumor as analyzed by (Rangayyan *et al.*, 1997), whereas Figure 2.8(c) and 2.8(d) show a benign and malignant tumor analyzed by (Guliato *et al.*, 2006).



(a) Benign (b) Malignant

(c) Benign (d) Malignant

Figure 2.8: Benign and malignant tumors. Tumors in (a) and (b) analyzed by Rangayyan *et al.* (1997). Tumors in (c) and (d) analyzed by Guliato *et al.* (2006)

Breast tumors and masses appear as dense regions in mammographic results and although there might be some exceptions to the general rule, a typical benign mass has a round, smooth, and well-circumscribed boundary, while a malignant mass has a spiculated, rough and blurry boundary (Varela *et al.*, 2006), (Cheng, *et al.*, 2006).

As it is difficult for radiologists to differentiate between benign and malignant tumors on mammograms, many recent studies have shown that techniques can be developed to assist radiologists to decide on the type of tumor by using quantitative methods. Guliato *et al.* (2006) derived a mathematical model for mammograms, in order to derive polygonal models of contours for an accurate classification of tumors. Rangayyan *et al.* (2000) developed a method to quantify the sharpness of the tumor boundaries. Varela *et al.* (2006) divided the tissue tumor border into three sections and analyzed these sections independently to decide if the mass lesion is benign or malignant by considering the shape of its interface shape. Kim and Min (2002) developed a mathematical model to count the number of jags of the breast tissue and breast tumor interface in order to differentiate between benign and malignant tumors. The most common and effective method for detecting breast tumors in their early stages is by performing mammogram screening. Mammography is currently the most effective modality used in breast cancer screening programmes (Sickles, 1997), (Anttinen *et al*., 1993), (De Koning *et al*., 1995), (Hendee *et al*., 1999), (Tabar *et al*., 1985), (Thurfjell *et al*., 1994), which is discussed in detail the following sections.

2.3.1 Mammography

At the current moment the modality of choice for breast cancer screening is mammography. Mammography is an X-ray technique developed specifically for the breast. It is based on the differential absorption of X-rays between breast tissue components such as the fat, connective tissue, tumor tissue and calcifications. Mammography is used both as a clinical tool to examine symptomatic patients and for screening purposes (Sickles, 1984). Requirements for mammography are high contrast, high spatial resolution, and minimal radiation exposure. High contrast is needed because differences in density between normal and pathologic structures of the breast tissue are small. The detection of MCCs requires both high contrast as well as a high spatial resolution. Minimal radiation exposure is essential as in screening programmes women frequently undergo mammography, often annually. Breast cancer can be recognized mammographically by the presence of a mass lesion (masses) or MCCs. The characteristics of mass lesions and MCCs are as follows (Burrel *et al*., 1996):

- **Mass Lesions**

  Most breast tumors, benign as well as malignant tumors are present as a focal mass lesion. The task of radiologists therefore is to discriminate between benign and malignant lesions. When a radiologist considers a lesion suspicious for containing a malignancy the patient will undergo additional examinations. The most important sign of malignancy is the presence of spiculation, which is a stellate pattern of lines directed towards the centre of the lesion. The border of a mass may also give information about the potential malignancy of a lesion. Benign masses are often characterized by sharp, circumscribed borders. Malignant masses on the

other hand have ill-defined or spiculated borders. The sharpness of the border however cannot be used as solitary criterion to identify malignancy, as some malignant masses, for example medullary carcinoma, colloid carcinoma and intracystic carcinoma, have circumscribed borders as well (Sickles, 1984). Moreover benign masses generally have poorly defined margins, for instance due to overlapping of the breast tissue or fibrosis. When a lesion is probably benign or when multiple similar masses are found in the breast the patient is often placed in a follow up protocol. Otherwise further examination is necessary to determine the nature of the mass (Sickles, 1986).

- **Microcalcifications**

  Another sign of malignancy is the presence of microcalcification clusters (MCCs). Microcalcifications develop in microscopically small cavities inside the lobuli or ducti. Microcalcifications inside the lobular unit are often due to benign conditions such as adenosis or fibro-adenoma (Popli, 2001). MCCs of ductal origin are more suspicious and may be the first sign of breast cancer. Intra-ductal microcalcifications can be diagnosed as benign or malignant by analyzing the shape of the cluster and the shape of the individual microcalcifications. Studies show that irregular, pleomorphic shapes of microcalcifications have a higher probability of being associated with malignant disease than those with round shapes and uniform size (Sickles, 1984).

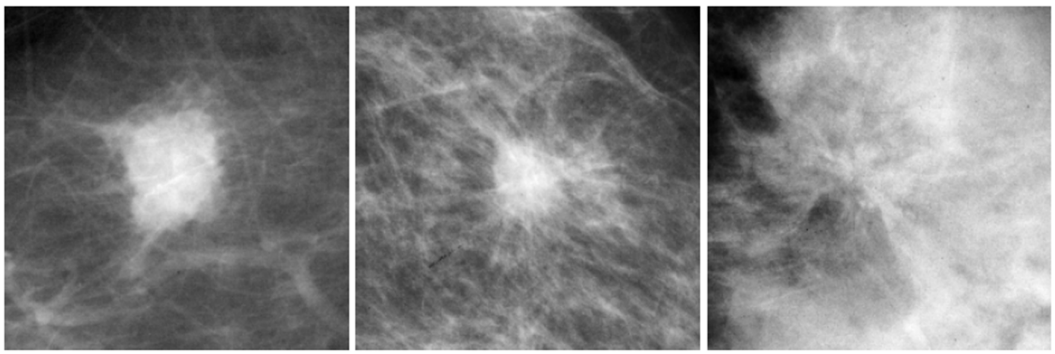
2.4 Screening for Breast Cancer

Early detection of breast tumors, especially breast cancer, can save thousands of lives each year. Screening is a very important step for early detection of diseases, which can locate breast cancers while they are still small in size and confined to the breast before they cause any symptoms. Screening is also important because breast cancers that cause discomfort to patients and that are big enough to be easily felt, tend to have already spread outside the breast to the other parts of the body (Monsees, 1995).

During breast cancer screening, the incorrect identification of a malignant abnormality as benign in breast cancer patients is generally referred to as a false positive (FP) (Kocur *et al*., 1996), (Fogel *et al*., 1998), as discussed in Section 3.3.6. The aim of breast cancer screening is early detection of breast cancers while keeping the number of false positive (FP) detections at a minimum. Similar to breast cancer screening, the performance of a medical diagnostic test is typically measured using the Receiver Operating Characteristic (ROC) curve analysis (see Section 3.3.6), where the four performance measures: true positive (TP), false positive (FP), true negative (TN) and false negative (FN) measure the *sensitivity* and *specificity* of the tested samples.

The earlier breast cancers are detected; the better the treatment options are available. A high patient recall rate, i.e. the percentage of mammographically screened women that is recalled for further assessment, generally improves the detection rate. This however, can lead to an increase in the number of FP detections resulting in unnecessary examinations (biopsies) and additional costs. Most countries have patient recall rates between 3 to 5 percent.

In many countries breast cancer screening programs using mammography have been started to detect cancers as early as possible. A screening program is defined as a program where an asymptotic group is invited to examine a specific disease on a regular basis. For breast cancer screening programs, only women are invited due to the very low incidence rate among men. A number of parameters must be chosen for a breast cancer screening program. The two parameters are: (i) the age range of women that are invited and, (ii) the time interval between two screening rounds. It is a highly debated subject at what age women should be invited for their first screening (Peer *et al.*, 1995), varying in practice between 40 and 50 years. Below the age of 40, the incidence rate of breast cancer is extremely small, increasing rapidly between the age of 40 and 50, which continues to increase more gradually for older women. The problem with screening young women is that their breasts contain much glandular tissue, yielding mammograms that are difficult to read due to denser tissues.



(a) (b) (c)

Figure 2.9: Examples of the most common signs of malignant abnormalities
(a) Circumscribed lesion (b) Stellate lesion (c) Architectural distortion

Breast cancers in young women are often aggressive and fast growing tumors, requiring short intervals between two screenings. After menopause, the breast

becomes less dense, making successful screening for small cancers more feasible. The upper limit of age for which women are invited for screening varies between 65 and 75 (Van Dijck *et al.*, 1997).

If the interval between two successive screening rounds is too large, a number of tumors that are detected in screening have already reached a stage with a lower chance of successful treatment. Tumors occurring during this interval are known as interval carcinomas. A large number of interval carcinomas may indicate that the screening interval should be made shorter. A short interval period will have a larger effect on the reduction of mortality, but is more expensive and women are exposed to a higher number of X-ray doses. In the United Kingdom, the screening interval is 3 years, a period that is considered too long by some researchers (Dean, 1996), whereas in Sweden and the Netherlands it is 2 years (Tabár *et al.*, 1997). However, in Malaysia, there is an absence of a national screening programme for breast cancer. As, breast cancer is the commonest cancer among Malaysian women, (18 percent of all cancer cases) (Lim *et al.*, 2008), age dependent screening should be performed by the NCR, i.e. 1 to 1.5 years for women below 50 years and 2 years for women over 50 years (Lim *et al.*, 2008), (Khuzi *et al.*, 2009).

In the United Kingdom only oblique films are used in screening, most other countries use both oblique and cranio-caudal films. The way mammograms are read also varies between countries. In some countries (for example the Netherlands) mammograms are examined by two radiologists, called double reading. Various approaches can be used to combine the findings of the two radiologists. Thurfjell *et al.* (1994) found that the sensitivity increases when double reading is practiced (when a case is recalled if either one of the radiologists

finds it suspicious), without changing the positive predictive value (PPV). In medical diagnostic tests using ROC curves, *sensitivity* represents the ratio of tumors which are marked and classified as tumor, to all marked tumors. *Specificity* represents the ratio of tumors which are not marked and also not classified as tumor, to all unmarked tumors, as indicated in Section 3.3.6. The study by Thurfjell *et al.* (1994) was criticized by Beam & Sullivan (1994) because the demonstrated increase in sensitivity is a mathematical fact; the question should be if the increase in sensitivity is worth the decrease in specificity. Another study has shown that double reading based on consensus between radiologists is a cost-effective screening procedure (Brown *et al.*, 1996).

Proving the efficacy of a screening program in a traditional epidemiological way is difficult due to the lack of an effective control group (Dean, 1996). If half of the population is offered screening, part of this group will not participate. Women in the group that is not offered screening and subsequently develop breast cancer typically do not seek medical attention until tumors are already in a late and incurable stage. If the group of non-participating women is large, a serious self-selection bias is present in the study. Even more important, women in the control group cannot be denied to have mammography on a regular basis. Women that are in a high risk group will especially request for mammography on a regular basis. Thus, if these high risk women are in the control group, it would then, reduce the number of cancers that are found at an incurable stage in the control group creating a biased experiment. This effect is called contamination and is a serious problem when the effect of screening is studied.

These factors make it difficult to prove a significant mortality reduction in screening. Comparing the number of breast cancer deaths with the number before screening is a common way to solve this problem, but it suffers from several drawbacks: the incidence of breast cancer may have changed, the treatment of breast cancer may have improved, or women may be more aware of abnormalities and seek medical assistance earlier than they might have done before. Another complicating factor is the long time it takes before a screening program reaches its maximal reduction in mortality.

Early screening programs were based on breast examination using palpation, either by the woman or a physician. No significant reduction in mortality has been reported on randomized trials using this type of screening (Newcomb *et al.*, 1991) although a few other studies suggest a small benefit (Baines, 1992). In 1963 the Health Insurance Plan (HIP) project was started in New York, the first large screening experiment using mammography as the main screening modality, together with palpation. A reduction in mortality was found for women in the group that underwent screening, a success that could be achieved because palpation and mammography were hardly practiced by women in the control group (Shapiro *et al.*, 1971). The success that was reported stimulated other countries like Sweden, Finland, the United Kingdom, Canada and the Netherlands to start experiments with breast cancer screening. The results of large numbers of randomized studies have been published by Dean (1996). It is commonly accepted that these studies show a reduction in mortality for women that take part in a screening program, especially for the age group between 50 and 70 years old. This is confirmed by other non-randomized and cohort studies in the United Kingdom and the Netherlands.

In a number of studies the possible benefit of inviting young women between 40 and 50 years old to a screening program was examined, but no unequivocal results were obtained. Some studies suggested a reduction in mortality (Thurfjell *et al.*, 1996); others did not find evidence for this (Peer *et al.*, 1995). It was shown in Van Dijck *et al.* (1997) that screening is beneficial at least until the age of 75. Due to the limited number of women over 75 that participated in screening, no significant results could be obtained for this age group.

An important trial of mammography screening was conducted between 1977 and 1984 in Sweden (Tabár *et al.* 1985). This trial concerned women aged 40 and older. The women were divided randomly into two groups, namely the: *study group* and *control group*. Each woman in the study group was offered screening every 2 or 3 years depending on their age, while women in the control group were not offered screening. The results obtained after seven years of follow up showed a 31 percent reduction in breast cancer mortality rate for the women in the study group who were invited for screening (Shapiro *et al.*, 1982).

Different trials were undertaken to determine whether these screening programmes achieved their goals. The eight most important trials are the following: Chu *et al.* (1988), Alexander *et al.* (1999), Bjurstam *et al.* (1997), Frisell *et al.* (1997), Tabár *et al.* (1995), Miller *et al.* (1992a, 1992b), Andersson *et al.* (1988), and Andersson & Janzon (1997). Most of these trials show a significant reduction in breast cancer mortality, especially for women aged between 50 to 70 years. These results have been used to guide screening programmes worldwide. The efficacy of screening mammography, especially for women in the age group from 40 to 49 years, remains controversial. Due to this reason, the American

Cancer Society (ACS) has recommend mammography screening on an annual basis for all women beginning at the age 40 (American Cancer Society, 2009).

2.4.1 Errors In Screening

Several studies have shown that approximately 20 percent of all interval carcinomas are visible on previously screened mammograms (Savage *et al.*, 1994), (Vitak, 1998), (Burrel *et al.*, 1996). Of all cancers detected during screening, 20 percent are retrospectively considered actionable on previously screened mammograms (Harvey *et al.*, 1993), (Bird *et al.*, 1992), (Van Dijck *et al.*, 1993). These numbers suggest that a considerable improvement in mortality reduction is possible if these errors could be prevented. When mammograms are examined retrospectively for signs of cancer, the abnormality is considered occult (i.e., nothing is visible on the previous mammogram) or classified as a minimal sign or screening error. An abnormality is called a minimal sign if something abnormal is found in the Region of Interest (ROI) that is not suspicious enough to recall. If signs of cancer (malignant tissue) are present that are actionable, it is called a screening error. However, many tumors do show clear signs of cancer on previously screened mammograms and many are found by automated detection systems at high specificity levels. A problem with this type of study is the subjective nature of the findings: normal, minimal sign and screening error, since the definitions of the findings vary considerably between radiologists.

There are two reasons as to why women with a visible tumor are not recalled for a follow up. The first possibility is that the sign was overlooked, and was not examined at all by the radiologist. The second possibility is that the sign was examined but it was considered benign, normal (no tumor), or not found

suspicious enough for further examination. So far, only a few studies have focused on the reasons as to why errors are made in the field of mammography (Hartswood *et al.*, 1998), (Hutt, 1996), although some work has been done in other medical areas (Friedman, 1999). Much work on the signal detection theory (Green & Swets, 1966) has been done in psychology departments, some related to the medical field (Laming, 1995). Psychophysical evidence exists that inserting extra abnormal signals to increase the target rate improves the performance when the target rate is very low, which is the case in breast cancer screening programs (Laming, 1995).

Mammographic signs of breast cancer missed in most screening programs are mass lesions and architectural distortions (Bird *et al.*, 1992), (Burrel *et al.*, 1996), (Vitak, 1998). Masses are often obscured by glandular tissues or have low contrast or no clear cancer (malignant) signs, like fuzzy edges or spicules. MCCs on the other hand are more easily detected by radiologists, but are often hard to classify between benign and malignant types.

2.5 Imaging Modalities

At the moment the modality of choice for breast cancer screening is digital mammography. For further examinations, or when digital mammography is not sufficient, other modalities are used, which include: ultrasonography (US) and magnetic resonance imaging (MRI). The following section gives an overview of different imaging modalities used for mammography screening.

2.5.1 Digital Mammography

Although most radiologists are more comfortable with the use of screen-film combinations, its disadvantages are obvious. Once an image is printed using screen-film technology, it can no longer be manipulated, and any information available in digital format, but not captured on the printed image will be lost. Furthermore screen-film combinations have limitations in detecting subtle soft tissue lesions, especially during the presence of dense glandular tissue (Kobatake *et al*., 1998).

To overcome the limitations of screen-film mammography, digital mammography was introduced (Lewin *et al*., 2001). Digital mammography provides several advantages over screen-film mammography such as the easy access to images, the use of computer-aided detection methods, improved means of transmission and retrieval and storage of images, and the use of a lower average dose of radiation without compromising the diagnostic accuracy.

In a recent study, the authors in Pisano *et al.* (2005) compared the diagnostic accuracy of digital and screen-film mammography. In this study a total of 49,528 asymptomatic women underwent both digital and film mammography. Breast cancer status was ascertained by a breast biopsy or a follow-up mammogram. This study showed that the overall diagnostic accuracy of digital and film mammography was similar, digital mammography however turned out to be more accurate in: women under the age of 50 years and women with radio-graphically dense breasts. As discussed in Section 1.1 previously, a notable advantage of digital mammography is the possibility of computerized detection of breast cancer.

2.5.2 Ultrasonography

The role of Ultrasonography (US) in breast imaging is a subject of ongoing discussion. Studies that have been performed using US as a mammogram screening tool failed to establish it's the efficiency of US. Thus, it has been concluded that US should not be used as a mammogram screening tool (Rahbar *et al.*, 1999).

2.5.3 Magnetic Resonance Imaging

High resolution Magnetic Resonance Imaging (MRI) of the breast has recently emerged as a sensitive instrument for the detection of breast cancer. MRI has proven to be useful in screening younger women with dense breasts who are at a high risk of developing breast cancer (Stoutjesdijk *et al.*, 2001). MRI can also be used as an adjunct to mammography for selected patients. However, MRI has a significant false positive (FP) rate (see Section 3.3.6) and it is not available in all areas due to being more expensive than digital mammography. Other limitations of MRI are that, it requires contrast injection and it can cause problems with claustrophobia. Thus, at the moment MRI remains limited to specific problem solving situations for patients at high risk for cancer.

2.6 Mammogram Analysis Using Digital Mammography

Digital mammography is an accepted and often preferred screening modality to detect breast abnormalities (Powell & Stelling, 1994). An X-ray passes through the breast, being absorbed selectively by different tissue types and emerges to be recorded onto a film or plate (Roebuck & Blamey, 1990). Screen-film mammography is generally the most common form today, where the X-rays strike a screen which emits photons that expose a photographic film (Dance, 1993). Over the years the radiation dose to the patient has decreased. Radiation exposure is

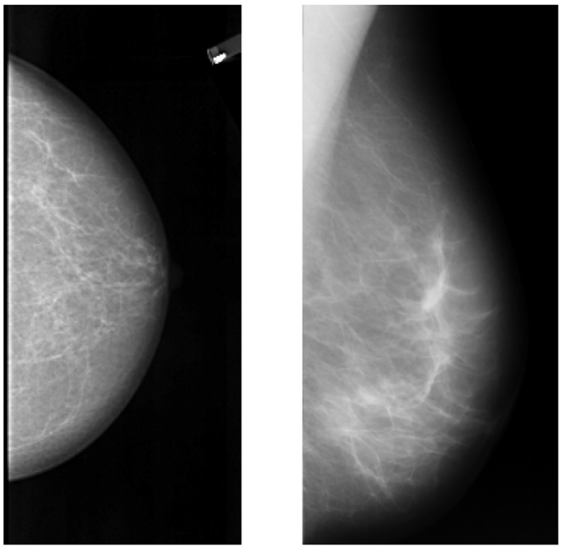
measured either in Roentgen (R) or Coulomb per kilogram (C/kg), where one $R = 2.58 \times 10^{-4}$ C per kg. The first dedicated mammography unit introduced commercially in 1969 typically delivered an 8 to 12R patient dose. By 1976 a screen-film system was introduced which lowered this dose to approximately 0.08R (Andolina *et al.*, 1992). At the same time as radiation dose decreased, image quality increased, through anti-scatter grids that absorb scattered X-rays and increase contrast of the image and provide improved compression systems and automatic exposure (Powell & Stelling, 1994).

There are alternatives to the screen-film mammography. Xeromammography, now an obsolete technique, was established when non-screen films were being used (Roebuck & Blamey, 1990). Just as the screen-film combination displaced Xeromammography, recent digital detectors have the potential to become the dominant technology due to the advantage of generating digital images; digital mammography.

2.6.1 Breast Positioning in Digital Mammography

In digital mammography, the breast is compressed between two parallel plates to spread the breast tissue and make the breast a block of uniform *physical thickness* for the X-rays to pass through. This compression can be performed at different angles to generate different orientations of the breast. The two standard views used in screening mammography are the: *Cranio-Caudal* (CC) view, generating a top to bottom view of the breast and the *Mediolateral Oblique* (MLO) view, a side-on view at approximately 45°. Examples of each view are shown in Figure 2.10 with the CC view shown in Figure 2.10(a) and MLO view in Figure 2.10(b).

While the breast is compressed to a uniform physical thickness during mammography, the *radiographic density* of each tissue type present in the breast determines the appearance of the mammogram. Radiographic density is the term used to describe the level of attenuation that the X-rays experience from the source to the detector. The higher the density, the less developed the film, resulting in appearance from fully exposed (black) to unexposed (white) depending on the tissue type. The fat in breasts has a low density, allowing the X-rays to pass through easily to expose the film, hence fatty areas of the mammogram are dark, in some places almost as dark as portions of the image where there is no tissue (background pixels). The glandular tissue in the breast has a somewhat higher density, resulting in brighter areas, as does the tissue of the pectoral muscle. Microcalcifications are very high in density; some lesions also have high density characteristics.



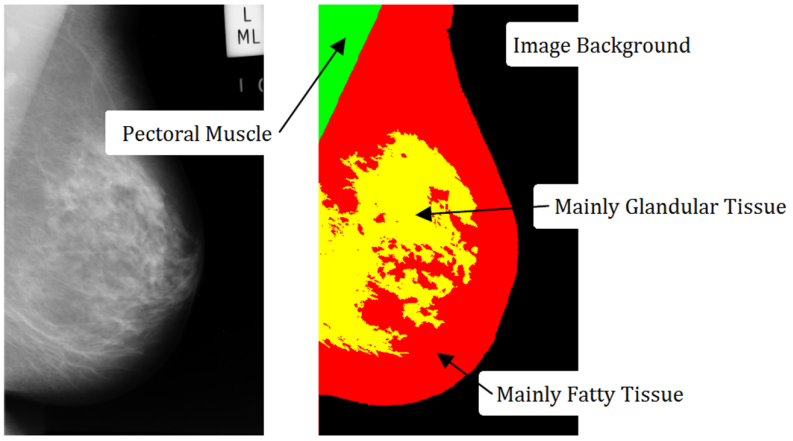
(a) (b)

Figure 2.10: The two standard views of the breast used in screening mammography (a) The craniocaudal view, from the head down (CC view) (b) A mediolateral oblique view, with the breast viewed from the side (MLO view)

As X-rays pass through a 3D breast to create a 2D image, the brightness of a pixel on the mammogram often represents the superposition of a number of tissue types that the X-ray has passed through on its way. Superposition presents several problems for segmentation of the breast tissue. It also presents problems with diagnosis. Microcalcifications are of much higher density than a comparable volume of glandular tissue. However, when superimposed on a large amount of glandular tissue, the 'bright' microcalcifications also appear less significant (Roebuck & Blamey, 1990).

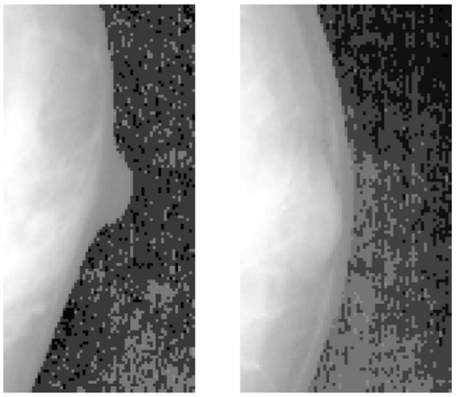
2.6.2 Breast Regions in Digital Mammograms

Regions present in the mediolateral oblique (MLO) view of a mammogram (see Figure 2.10(b)) are now identified to give the reader a chance to understand what part of the image is being discussed when certain names are used. The simplest distinction is to differentiate the image between breast and non-breast regions. The non-breast regions include the image background, labels, scanning artifacts and tapes, which might be superimposed over the breast region.



(a) (b)

Figure 2.11: Mammogram decomposition (a) Original mammogram image (b) Attempted decomposition of mammogram in (a) into separate breast regions



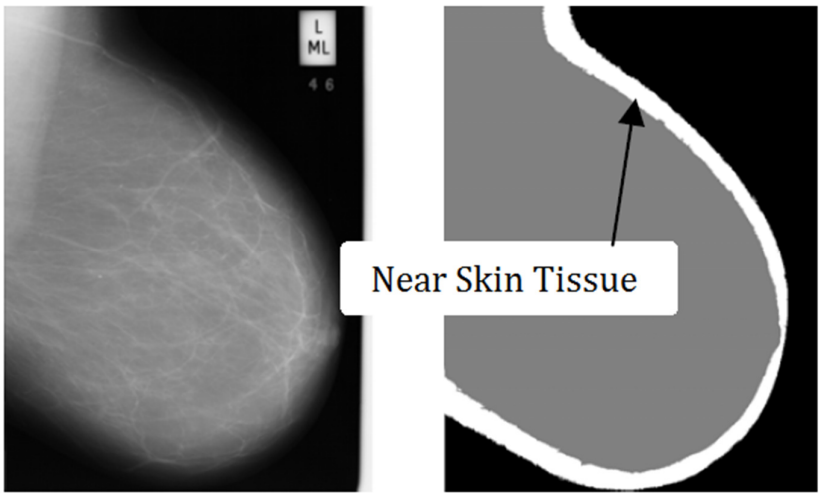
(a) (b)

Figure 2.12: Magnified views showing breast nipple (a) Nipple in the breast profile (b) Breast profile without the nipple

While non-breast regions are straight forward to identify, the division between regions inside the breast is less distinct due to the overlap of tissue types. An attempt is made in Figure 2.11 to indicate different breast tissues. The mammogram can be divided into fatty tissue (dark and glandular tissue), which is high in pixel intensity. The pectoral muscle is a characteristic feature of the MLO view of a mammogram and presents itself as a bright triangle in the top left or right corner of the mammogram (Kwok *et al*., 2004).

Another feature is the nipple which may not necessarily be seen in the breast profile in Figure 2.11. A close-up image with the nipple in the breast profile and breast profile without a nipple is shown in Figure 2.12. Near the border of the breast lies a region of fatty tissue termed as the near-skin tissue. This region results due to the poor compression of the breast at the edge causing a gradual decrease in thickness towards the skin-line. The near-skin tissue region is not usually visible in an original mammogram image without contrast enhancement. Figure 2.13 shows an original mammogram image and an enhanced version

(mask) of the original image indicating the near-skin tissue boundary of the breast profile (Wirth *et al.*, 2007).



(a) (b)

Figure 2.13: Near-skin tissue of a mammogram barely visible in mammogram
(a) Original mammogram image
(b) Enhanced image indicating the near-skin tissue as the mask of (a)

2.6.3 Types of Breast Tissues

The appearance of the breast tissue in a mammogram varies between images. The process of involution leads to the change from a predominantly bright, glandular tissue filled image in younger women to a darker, mostly fatty image (Roebuck & Blamey, 1990). Breasts can be divided into a number of classes based on the appearance of the glandular tissue (Wolfe, 1976). Breast tissues are classified into three major types based on their density: (i) fatty, (ii) fatty-glandular, and (iii) dense-glandular. An example of the different densities of glandular breast tissues is shown in Figure 2.14.

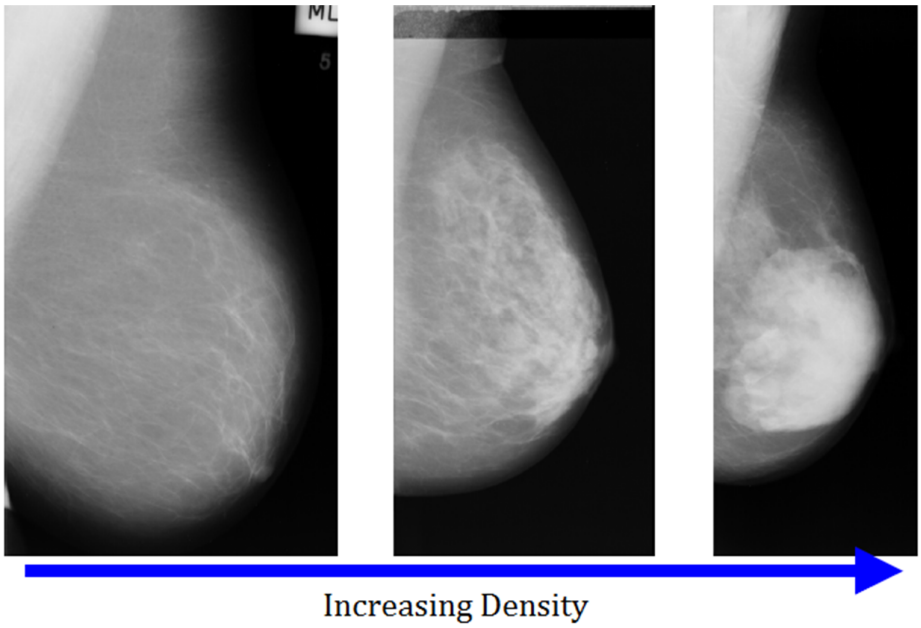


Figure 2.14: Three mammograms images with different breast tissue densities. From left to right: fatty, fatty-glandular and dense-glandular.

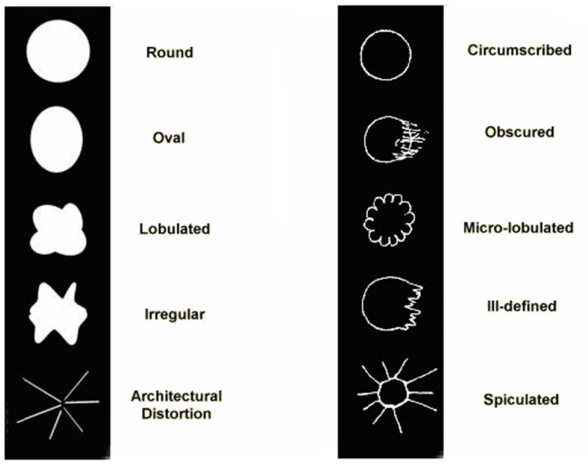
2.6.4 Present Clinical Protocol

This section presents the clinical protocol followed by radiologists for mammographic examination and interpretation. Standardized mammographic interpretations follow from the *Breast Imaging Reporting and Data System* (BI-RADS) lexicon. The American College of Radiology (ACR) developed the BI-RADS as a measure for mammographic interpretation for radiologists. The BI-RADS provides a mechanism for describing the characteristics of a given abnormality including the final pre-pathology finding (BI-RADS, 2010).

For classification of mass lesions the borders, shape and relative intensities of the lesions are important descriptive features. In the following subsections, the relevant BI-RADS descriptors and assessment categories are presented.

2.6.4.1 BI-RADS Descriptors and Assessment

BI-RADS descriptors are important factors for predicting malignancies that are assessed and provided by radiologists. Mass narratives include the overall shape description, the border region margin regularity and the relative intensity of the mass region compared with the ambient normal tissue intensity. The BI-RADS lexicon provides a four-category rating for assessing the overall breast tissue characteristics in terms of fibro-glandular composition (BI-RADS, 2010). The composition categories relate to the degree of interpretation difficulty. Similarly, the BI-RADS gives a five-point overall assessment that is related to the degree of probable malignancy or necessary follow up work.



(a) (b)

Figure 2.15: BI-RADS mass descriptors for (a) shape (b) margin

2.6.4.1.1 BI-RADS Mass Descriptors

BI-RADS have established mass descriptors such as for shape and margin, for the detection of mass lesions as indicated in Figure 2.15. The shape and margin properties of the BI-RADS mass descriptors are as follows (BI-RADS, 2010):

- *Shape:* The shape of the mass is described with a five-point assessment: round, oval, lobular, irregular and architecturally distorted as shown in Figure 2.15(a).

- *Margin:* The mass margins modify the boundaries. For example the overall shape of the mass may be round, but close inspection may reveal scalloping along the border, which may indicate a degree of irregularity or a lobular characteristic. The margins are rated with a five-point system: circumscribed (well-defined/sharply-defined) margins, obscured margins, micro-lobulated margins, ill-defined margins and spiculated margins as shown in Figure 2.15(b).

2.6.4.1.2 BI-RADS Assessment Categories

The BI-RADS assessment categories are defined for standardized interpretations of mammographic findings. Each category provides the overall assessment related to the findings and the necessary follow up required. The assessment categories are summarized in Table 2.1 (BI-RADS, 2010).

2.6.5 Mammogram Interpretation

Radiologists interpret mammographic examination in the form of a mammogram report. A mammogram report describes the findings i.e. breast abnormalities, and provides the radiologist's impression based on BI-RADS with recommendations on the appropriate actions to be taken**.** The elements of the mammogram reporting from BI-RADS are shown in Table 2.2 (BI-RADS, 2010).

Table 2.1: BI-RADS assessment categories (BI-RADS, 2010)

Category	Assessment and Recommendation	Findings
Category 0	Incomplete Assessment (No recommendation)	Needs additional imaging evaluation.
Category 1	Negative (No action required)	The breasts are symmetrical and no abnormalities are present.
Category 2	Benign Finding (No action required)	This is a negative mammogram and no abnormalities are present.
Category 3	Probably Benign Finding (Short interval follow-up suggested)	A finding has a high probability of being benign.
Category 4	Suspicious Abnormalities (Biopsy should be considered)	These are lesions that do not have the characteristic morphologies of breast cancer but have a definite probability of being malignant.
Category 5	High Suggestive of Malignancy (Appropriate action should be taken)	These lesions have a high probability of being malignant.

2.7 Summary

This chapter discussed breast cancer detection and digital mammography. General information about the structure and functions of the breast, breast tumors and literature regarding breast cancer screening were presented at first. Next, background issues concerning different imaging modalities such as digital mammography, ultrasonography (US) and magnetic resonance imaging (MRI) were reviewed and discussed. Towards the end of this chapter, the analysis and interpretation of digital mammograms using the BI-RADS lexicon was discussed.

Table 2.2: Elements of mammogram reporting from BI-RADS

Mammogram Element	Description
Findings	Breast abnormalities (i.e. mass lesions and MCCs) found from mammograms in terms of size, location and shape characteristics. • *Primary signs of breast cancer*: spiculated masses and clustered pleomorphic MCCs. • *Secondary signs of breast cancer*: asymmetrical tissue density, skin thickening, retraction and focal distortion of tissue.
Impression	Contains the radiologist's overall assessments (findings and breast abnormalities) using the BI-RADS lexicon.
Recommendation	Depending on the assessment, the recommendation contains specific instructions on what actions should be taken. For example, in Table 2.2, the radiologist could recommend: • For Category 0: Additional imaging such as spot views, MRI etc. • For Category 1 and 2: No action is necessary. • For Category 3: A six months follow-up procedure is required to establish the finding's stability. • For Category 4 and 5: A biopsy is required.

CHAPTER 3

ELEMENTS OF COMPUTER-AIDED DETECTION

3.0 Overview

This chapter presents and discusses the background literature on the computer processing of digital mammograms. In Section 3.1, computer-aided detection systems are introduced with the literature review of computerized breast cancer detection techniques presented in Section 3.2. Section 3.3 highlights and identifies the key techniques and algorithms used in this research to develop a framework for the computerized detection of breast cancer. Section 3.4 presents the fundamental concepts of digital image processing with emphasis on image segmentation techniques used in digital mammography applications. Lastly, Section 3.5 emphasizes on the use of texture-based analysis for the purpose of feature extraction in pattern classification problems.

3.1 Computer-Aided Detection Systems

In recent years major effort has been made to develop digital mammography applications which can assist radiologists in the detection and characterization of malignant and benign abnormalities. One such application of digital mammography is *computer-aided detection*, as discussed in Section 1.1 previously.

Computer-aided systems identify and mark suspicious regions on mammograms to bring them to the attention of radiologists. These systems minimize search, perception and interpretation errors even if radiologists fail to recognize

suspicious abnormalities. Computer-aided detection is intended to be used after the radiologist has completed evaluation of the mammographic images and has made an initial decision whether patient recall is required (Hutt, 1996). As an example, if the radiologist identifies an abnormal region on a mammogram during initial reading and that area does not get marked by the computer-aided method, the radiologist is advised to interpret the mammogram as positive and to recall the patient for further work up. Since, computer-aided detection is proposed as an adjunct to digital mammography to decrease search, detection and interpretation errors (see Section 1.1), the radiologist makes the final decision if a clinically significant abnormality exists and decides whether further diagnostic evaluation is warranted (Jirari, 2005).

The hope lies in the fact that computer-aided detection systems will improve the sensitivity of digital mammography without substantially increasing patient recall rates (Bozek *et al.*, 2008). The following section provides the review of computerized detection of breast cancer (mass lesions and MCCs) in digital mammography applications.

3.2. Review of Computerized Breast Cancer Detection Techniques

As it is known, the goal of computerized breast cancer detection in digital mammography is to identify the presence of abnormalities such as mass lesions and MCCs. Lesion detection is possible from a single mammogram image, as is the detection of MCCs, which is subjected to a wide volume of publications. Whichever way they are detected, masses and MCCs need to be classified into their malignant and benign types; a technique many authors have tried to automate.

Many papers have been published on enhancing digital mammograms for optimal viewing, mainly for computer-aided detection of MCCs (Mutihac *et al.,* 1998), (Strickland *et al.,* 1996) and (Cernadas *et al*., 1996) and masses (Woods & Bowyer, 1996). Most of these studies provide evidence that the radiologists perform better on computer enhanced images. Important work was done by Aylward *et al.* (1998) and Netsch *et al*. (1998) to transform mammograms in such a way that they can be printed or examined on a monitor optimally. For example, the dark area near the skin line can be enhanced (Karssemeijer & te Brake, 1996), (Bynd *et al*., 1997) and the pectoral muscle can be filtered out (Nicolaou *et al.*, 2008), largely reducing the intensity range in the mammogram. Good contrast will be available in the whole Region of Interest (ROI), both in the pectoral area as well as near the skin line. Furthermore, the authors in Highnam *et al.* (1996) described a method to filter scatter from digital mammograms.

Both mass and MCC processing follows a similar set of *mammogram preprocessing* steps. At first, the mammogram images are enhanced to highlight some property of the desired regions (ROIs) through either spatial filtering or time-frequency based methods (Thangavel and Karnan, 2005b). Heuristic features are then computed from the enhanced images (ROIs) and some basic classification is performed to differentiate between masses (Woods & Bowyer, 1996) and MCCs (Bazzani *et al.*, 2001), (Lu & Bottena, 2001). Further classification can be performed to identify masses and MCCs into malignant and benign types (Veldkamp *et al*., 2000), (Rahbar *et al*., 1999), (Jiang *et al*., 1998), typically using machine learning frameworks such as ANNs, statistical classifiers such as SVMs, or some sort of decision tree mechanism. During classification, features such as: shape, size, and

texture properties (such as statistical distribution of regions and other measures of texture) of the ROIs should be taken in account (Martì *et al*., 2003).

The search for abnormalities (masses and MCCs) in mammogram images generally uses the breast profile boundary to constrain processing to the breast area only, avoiding spending time processing non-breast regions. In order to constraint mammogram processing to the breast tissue area only, *mammogram segmentation* is a fundamental step to suppress un-important regions in mammograms (Wirth *et al*., 2007). Knowledge of the pectoral muscle may also be used in single image analysis, referred to as *bilateral comparison*. As breast tumors are frequently projected in the lower areas of the pectoral region and the presence of the intensity gradient at the pectoral muscle edge may easily generate false alarms or miss tumors (Karssemeijer & te Brake, 1998), thus the location of the pectoral muscle may be used to subtract its contributing intensity from the image.

Many attempts have been made to identify mass lesions and MCCs (Cascio *et al*., 2008), (Domínguez & Nandi, 2008), (Li *et al*., 1997), (Song *et al*., 2009), (Wirth *et al*., 2007), (Martì *et al*., 2003) in order to classify between malignant and benign types. Approaches to mass and MCC detection have been based on concepts which mainly include: template matching (Özekes *et al*., 2005), wavelets (Soltanian-Zadeh *et al*., 2004), (Mousa *et al*., 2005), (Gorgel *et al*., 2009), and measures of texture (Varela *et al*., 2001), (Oliver *et al*., 2007), (Mudigonda *et al*., 2001), (Martì *et al*., 2003), (Lyra *et al*., 2008), (Karahaliou *et al*., 2008), (Bovis & Singh, 2002). The list of methods used is extensive, only recent approaches for detection of mass lesions and microcalcifications/MCCs are discussed in the following section.

3.2.1 Detection of Microcalcifications/MCCs

The detection of microcalcifications is an important topic in computerized detection of breast cancer, because it is a task that radiologists also find challenging. To achieve a good positive predictive value (PPV) it is important to be able to discriminate between malignant and benign abnormalities, because only 20 percent of all microcalcification clusters (MCCs) are due to malignant processes. Pointing out all microcalcifications is a tedious and time consuming task, not suited for use in clinical practice. So far only few completely automated methods have been published (Sorantin *et al.*, 1998).

The authors in Kaufmann *et al*. (2001) defined a MCC as three or more microcalcifications within a 1cm diameter circle with cluster features such as: calcification number, cluster area, and statistical measures of inter classification distance combined with individual calcification features. These features were used to classify MCCs into benign and malignant categories by evaluating the k-nearest neighbor and Bayesian classifiers (Kaufmann *et al.*, 2001). Similar sets of MCCs features have been used by Jiang *et al.* (1998). The authors in Jiang *et al.* (1996) developed a method that outperformed five radiologists, using an ANN that classified MCCs based on eight features that were computed for each cluster. Bottema and Slavotinek (2001) determined the convex hull of a MCC by using the ratio of the maximum and minimum distances from the convex hull boundary to its geometric centre in order to separate clusters of large DCIS calcification from benign calcifications.

Microcalcifications in digital mammograms represent a sharp transition in intensity as they are quite bright; hence enhancing mammograms to highlight

small objects representing high frequency is used by most approaches (Soltanian-Zadeh *et al*., 2004), (Mousa *et al*., 2005), (Gorgel *et al*., 2009). The use of wavelets is a popular method to extract high frequency regions from mammogram images in order to search for microcalcifications and MCCs (Wang & Karayiannis, 1998), (Brown *et al*., 1998), (Bazzani *et al*., 2001), (Lu & Bottena, 2001). Other techniques include: Markov random field models (Karssemeijer, 1992) (Veldkamp & Karssemeijer, 1998), Spatial filters (Diahi *et al*., 1998), (Gürcan *et al*., 1998), Box-rim filters (Bottema & Slavotinek, 1998) and background subtraction with a model of the background produced by methods as Gaussian blurring (Lu & Bottena, 2001) or polynomial modeling (Bottema & Slavotinek, 1998), (Lu & Bottena, 2001).

Dealing with noise in mammograms is important for MCC detection algorithms. Most methods use local adaptive thresholding, because noise levels vary across mammogram images. Nishikawa *et al*. (1994) used an initial global threshold level, followed by a locally adaptive threshold step. The authors in Chitre *et al.* (1994) computed a local threshold image and used the local deviation of grey levels as a threshold to decide whether or not a pixel belonged to a MCC. Karssemeijer (1992) developed a method for this purpose which was improved by Veldkamp & Karssemeijer (1998). The advantage of a global correction approach is that the statistics are much better than for local estimations of the appropriate threshold level. Several other research groups such as Mutihac *et al*. (1998) and Strickland & Hahn (1996a) focused on noise equalization.

The properties of individual microcalcifications are important for their grouping. A large percentage of mammograms contain benign calcifications (Bottema *et al*., 2001). However, the more calcifications there are per unit area, the more likely is

the abnormality to be malignant (Roebuck & Blamey, 1990). Automated segmentation of microcalcifications can lead to complicated algorithms using more than one of the methods outlined above. As an example, the authors in Doi *et al.* (1993) and Yoshida *et al.* (1994) followed a series of steps which included: wavelet transform based processing, global thresholding, morphological erosion, local thresholding, texture analysis and clustering. As in the above example, using a variety of operations leads to a number of parameters leading to the problem of tuning due to interdependence of the parameters. However, adaptive tuning of parameters can be used to improve the success rate of the algorithm while keeping the False Positives (FPs) less. During breast cancer screening, the incorrect identification of a malignant abnormality as benign in breast cancer patients generally is referred to as a false positive (FP) (Kocur *et al*., 1996), (Fogel *et al*., 1998) (see Section 3.3.6). Similar to breast cancer screening, the performance of a medical diagnostic test in digital mammography applications is typically measured using the Receiver Operating Characteristic (ROC) curve analysis as presented in Section 3.3.6, where the four performance measures: true positive (TP), false positive (FP), true negative (TN) and false negative (FN) measure the *sensitivity* and *specificity* of the tested samples.

The authors in Anastasio *et al*., (1998) used Genetic Algorithm (GA) to tune the parameters in the case of the algorithm of Doi *et al.* (1993), which led to a sensitivity increase from 80 to 87 percent with a FP rate of 1 per image. Kobatake *et al*. (1998) used subtraction with several smoothed images, each produced with a top-hat transformation (using *morphological operations*) using several structuring elements (STRELs) to increase sensitivity and to remove the False Positives (FPs) causing 'elongated shadows' due to glands and blood vessels in the breast tissue.

The authors in Highnam and Brady (1999) proposed another way to remove FPs, is by performing direct segmentation of the curve-linear structures in a mammogram. Other methods to segment potential microcalcifications and MCCs include: mathematical morphology (Zhao *et al*., 1992), (Dengler *et al*., 1993), (Hagihara *et al*., 2001), (Kaufmann *et al*., 2001), (Bruynooghe, 2001), directional recursive median filtering (Cernadas *et al*., 1998), Fuzzy logic (Cheng *et al*., 1998), and fractal theory (Lee *et al.*, 2000), (Li *et al.,* 1997).

Recent approaches for identification of MCCs typically use features such as: shape, size and texture based properties (Sorantin *et al.*, 1998), (Meersman *et al.*, 1998), (Lu & Bottema, 2001), (Brown *et al.*, 1998), (Martí *et al.*, 2001), (Jiang *et al.*, 1998) applicable for pattern classification. The authors in Lee *et al.* (2001) compared four methods of microcalcification detection: (i) Karssemeijer's Markov random field (Karssemeijer, 1992), (ii) Strickland's wavelet based algorithm (Strickland, 1996), (iii) Grey-level "isophote" contours (Guillemet *et al*., 1996), and, (iv) Adaptive threshold based method (Wallet *et al*., 1997). The algorithms were tested on images from two databases, with better results for one database. The criterion for success was detecting at least 80 percent of the MCCs with as little FPs as possible. The algorithms by Wallet *et al*. (1997) and Guillemet *et al*. (1996) did not meet the 80 percent sensitivity requirement. In conclusion to the experiment performed by Lee *et al.* (2001), Karssemeijer's algorithm (Karssemeijer, 1992) is favored more as it produces less FPs than Strickland's wavelet based algorithm (Strickland, 1996).

3.2.2 Detection of Mass Lesions

A large variety of techniques have been applied to the problem of mass lesion detection, but most follow a two-step scheme that was described by Woods and

Bowyer (1996). First, one or more features are computed for each pixel, after which each pixel is classified and the suspicious pixels are grouped into a number of suspicious regions. In the second step, these regions are classified as normal or abnormal regions, based on regional features like size, shape, contrast and texture properties (Woods and Bowyer, 1996). Two signs can indicate the presence of a lesion: (i) a radiating pattern of spicules, or (ii) a central mass. To detect the whole range from architectural distortions to circumscribed mass lesions, both signs must be detected (Zhang & Giger, 1995), (Cascio *et al*., 2008).

3.2.2.1 Central Mass

The central mass of a lesion is a circular bright region with a diameter between 5mm and 5cm. Convolution of mammogram images with a zero-mean filter with a positive center and a negative surrounding area was used by a number of research groups to detect mass lesions (Sahiner *et al.*, 1996), (Zheng *et al.*, 1995), using the Laplacian of the Gaussian (LoG) and Difference of Gaussian (DoG) filters. This is an easy and intuitive approach to detect bright blobs, but may not be suited to find masses with lower contrast. Other approaches that are less dependent of the contrast are more useful, like template matching, a method used in some early research studies for the detection of the central mass (Ng & Bischof, 1992), (Lai *et al.*, 1989). In template matching, a model is made of the appearance of a mass and the mammogram is searched for regions that resemble this model. This approach is more related to the shape of the region, rather than the contrast. It is especially hard to detect low contrast masses using this method, however it may outperform convolution based approaches.

Other approaches for lesion detection focus on the analysis of the gradient patterns in the region of interest (ROI). The appearance of masses in mammograms varies and therefore the rigid approach of template matching is not very successful. In an area with a central mass, the orientation of the gradients will be towards the center of the mass. Statistical analysis of this type of pattern can be used to discriminate masses from other structures. The authors in Groshong and Kegelmeyer (1996) used a generalized Hough transform for detection of the central mass. The strongest edges in the ROI are accumulated in a Hough space where each location relates to a center and a radius. Masses yield peaks in this space (Groshong & Kegelmeyer, 1996). The authors in Zwiggelaar *et al*. (1999) applied a one-dimensional recursive median filter over a number of different angles to each pixel in order to detect the central mass.

Typically masses look very much similar to normal glandular structures, and they are only detectable due to asymmetry between the left and right breasts. Matching two breasts is a complicated procedure because there is only an approximate correspondence between the normal tissue in the left and right breasts and due to variations in compression and positioning, the variation in appearance is even made larger. Yin *et al.* (1991, 1993) applied a simple rigid body transform to align the skin line of the two (left and right) breasts. Another sophisticated approach includes matching corresponding points between the two breasts. Lau *et al.* (1991) use a set of 3 control points and an estimation of the nipple. Sallam and Bowyer (1994) used a more general warping method to match automatically detected landmarks of the glandular tissue. When the two breasts are correctly matched, subtraction and smoothing can be done to find a number of suspicious regions (Lau & Bischof, 1991), (Yin *et al.*, 1991), (Yin *et al.*, 1993).

More recent approaches towards detection of lesions focuses on texture based analysis (Varela *et al*., 2001), (Soltanian-Zadeh *et al*., 2003). Haralick's texture descriptors (Haralick *et al*., 1973), (Haralick, 1979) and Laws' texture measures (Laws, 1980) have been used quite extensively to detect mass lesions. Kegelmeyer Jr. (1994) used Law's texture measures together with the Analysis of Local Oriented Edges (ALOE) as features to determine the probability of abnormality for each pixel using a binary decision tree. Polakowski *et al*. (1997) also made use of Law's texture measures using a system based on several modules enhancing masses through a Difference of Gaussian (DoG) filter and calculating Law's texture measure along with features based on shape, size, and contrast with an ANN for classification of regions as malignant or benign. In more recent times, Khuzi *et al*. (2009) applied Haralick's texture descriptors for the identification of masses in digital mammograms. It has been reported that the success of lesion detection using Haralick's texture descriptors is highly dependent upon mammogram preprocessing (Khuzi *et al*., 2009)

3.2.2.2 Spicules

Architectural distortions are a straight forward indication of a malignant lesion. When a mass is surrounded by spicules, it is likely to be malignant. Many stellate lesions are easier to detect by their spicules than by their central mass. Kegelmeyer *et al*. (1994) computed histograms of local gradient orientations, which indicated that areas with a spicule pattern have flatter histograms than normal areas. This feature was combined with four texture features; however, the good results of Kegelmeyer *et al*. (1994) could not be reproduced by other groups (Woods & Bowyer, 1996). Analysis of texture in the Hough space was the basis of an approach proposed by (Zhang and Giger, 1995). The authors in Parr *et al.*

(1996a, 1996b) developed a model for detection of spicules using the Principle Component Axis (PCA) and achieved good results for the detection of spicules. Other techniques include pixel-level algorithms that detect both spicules and masses, but are very sensitive and are known to signal many false positives (FPs) (see Section 3.3.6).

3.2.2.3 Normal and Abnormal Regions

The classification of normal (malignant) and abnormal (benign) regions is a well-studied subject in digital mammography (Veldkamp *et al*., 2000), (Rahbar *et al*., 1999), (Jiang *et al*., 1998). Articles reviewed on this topic focus on edge analysis of masses, where vague or spiculated edges indicate malignancy, and sharp and well defined contours are likely to belong to a benign abnormality (Martì *et al*., 2003). The authors in Rangayyan *et al.* (1997) indicated that radiologists outlining lesions may induce a bias because the way they outline spicules or vague regions can be incorporated in the computed lesion features. Other interesting work was done by Giger *et al.* (1994) and Hui *et al.* (1995) who used a radial edge gradient method to discriminate malignant and benign lesions. Pohlman *et al*. (1996) developed a feature describing the tumor boundary roughness for differentiation between benign and malignant masses.

A number of articles have been published on the topic of discriminating real lesions from suspiciously looking normal tissue (Sahiner *et al*., 1996), (Wei *et al*., 1995), (Polakowski *et al*., 1997), (Yin *et al*., 1993). In most cases, features or heuristics are computed over a large region containing the suspicious area (Wei *et al*., 1995), however, most research groups have segmented the suspicious area to reduce the number of FPs. (Domínguez and Nandi, 2009), (Martì *et al*., 2003).

Segmentation of the suspicious breast tissue is useful in separating the abnormal breast tissue from the breast normal tissue, as it enables computation of features related to the edge of the region, as well as other properties such as the contrast, shape and size. The segmented suspicious area is typically known as the Region of Interest (ROI). Recent approaches by authors in (Khuzi *et al.*, 2009) and (Veldkamp *et al*., 2000) signify that using texture based features (Haralick *et al*., 1973), (Haralick, 1979), (Laws, 1980) a considerable improvement in classification between malignant and benign types is generally achieved by the removal of FP signals.

3.3 Computerized Detection of Breast Cancer

The goal of computerized breast cancer detection systems is to reduce the number of false positive (FPs) and to achieve high sensitivity for detecting cancers that radiologists might miss (Hutt, 1996). During breast cancer screening, the incorrect identification of malignant lesion as benign in breast cancer patients, generally referred to as a false positive (FP) (Kocur *et al*., 1996), (Fogel *et al*., 1998) as discussed in Section 3.3.6. The clinical utility of computerized systems depends upon the number of FPs per image, since radiologists must take extra time and care to inspect areas of the mammograms with FPs (Nagel *et al.*, 1998).

Computer processing of mammograms in digital mammography applications typically involves *image processing*. Similarly, for pattern classification between malignant and benign abnormalities in mammograms, classification (machine learning) is required. Figure 3.1 shows a typical framework of a computerized breast cancer detection system for digital mammography applications. The

following sections discuss in detail each of the six stages in the computerized detection of breast cancer (Hutt, 1996).

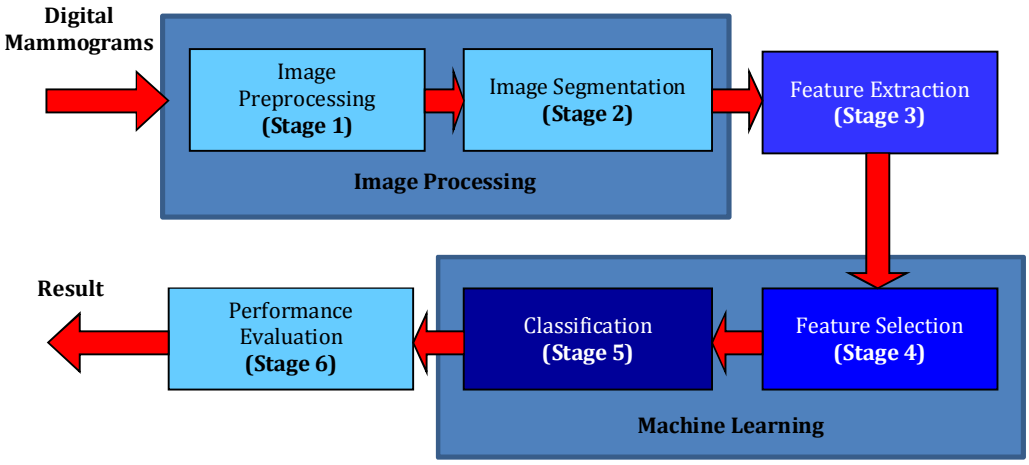


Figure 3.1: General framework the computerized detection of breast cancer (Hutt, 1996)

3.3.1 Image Preprocessing

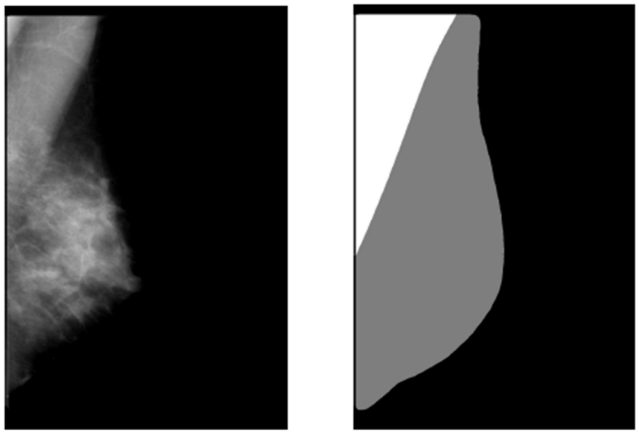
Preprocessing of digital mammograms generally involves noise and radiopaque artifact suppression and image adjustment. Image adjustment is usually achieved by performing contrast enhancement using image histogram measures. Increasing the contrast is very essential in mammograms, especially for dense breasts (Wang & Karayiannis, 1998). Contrast between the malignant and benign abnormalities maybe present in mammograms but may not be discernable to the human eye. As a result, differentiating between malignant and benign abnormalities is difficult (Wang & Karayiannis, 1998). Conventional image processing techniques may not work well on mammography images because of the large variation of feature sizes and shapes (Morrow *et al.*, 1992). There are a few approaches to enhancing mammographic features, as reviewed by the authors in Thangavel *et al.* (2005a).

One technique is to suppress background and digitization noises and the other is to increase the contrast of suspicious areas (Thangavel & Karnan, 2005b).

Noises due to intrinsic characteristics of an imaging device or imaging process will impact the sensitivity of the classification result. Several types of filters have been reported by Qian *et al.* (1994) in order to reduce imaging and digitization noises. Methods like straight line windowing (Chan *et al.*, 1987) and hexagonal windows (Glatt *et al.*, 1992) have been introduced as non-linear filtering techniques. Though non-linear filtering techniques have shown to be more successful for noise suppression than linear approaches, they do not necessarily show significant improvements in image detail preservation. Other techniques include Median filtering and edge-preserving. Median filtering locally adapts to grayscale images using an empirically derived threshold criteria (Lai *et al.*, 1989) and (Thangavel & Karnan, 2005b). Other techniques include edge-preserving smoothing, which searches for a homogeneous neighborhood in different directions of a given pixel and averages this neighborhood.

Mammographic artifacts are small emulsion continuity faults on X-ray mammogram films, which look like MCCs. Artifacts in mammograms normally contain the form of labels, markers and wedges in the unexposed air-background (non-breast) region. Such artifacts are usually radiopaque in the sense that they are not transparent to radiation. These artifacts are usually sharply defined and brighter than the microcalcifications, which are normally present in the background region of mammograms. One of the problems with the precise segmentation of the breast profile in mammograms is due to the existence of artifacts (e.g. a label overlapping the breast region), which often results in a non-

uniform background region causing many segmentation algorithms to fail. A robust artifact suppression algorithm based on area morphology developed by Wirth *et al.* (2004) removes radiopaque artifacts from the background region of mammograms (Yapa & Harada, 2008), (Wirth *et al.,* 2007). Thus, the approach proposed in this research applies the 2D median filtering (Thangavel & Karnan, 2005b) for the purpose of noise removal and the algorithm based on area morphology (Wirth *et al.*, 2004) is used for the purpose of radiopaque artifact suppression.



(a) (b)

Figure 3.2: Digital mammogram (a) Original mammogram (b) Segmentation of the mammogram into the breast area (grey) including the pectoral muscle (white) and background region (black)

3.3.2 Image Segmentation

Following noise suppression and artifact suppression, *image segmentation* is used to identify suspicious areas (ROIs) in the mammogram images. The aim of segmentation is to obtain ROIs containing all breast abnormalities from the breast tissue and locate suspicious lesions and MCCs from the ROIs. Mammographic abnormalities such as mass lesions are extremely difficult to identify because their

radiographic and morphological characteristics resemble those of normal breast tissues. Since a digital mammogram is a projection image, mass lesions do not appear as isolated densities but as overlaid over parenchymal tissue patterns.

During the last decade a number of mammogram segmentation techniques have emerged. These segmentation techniques include: Region growing (Kupinski & Giger, 1998), (Petrick *et al*., 1999), (te Brake & Karssemeijer, 2001), Markov Random Fields (Li *et al*., 1995), Fractal modeling (Lefebvre *et al*., 1995), (Li *et al*., 1997), Tree structured wavelet transform (Qian *et al*., 1995), Adaptive density-weighted contrast enhancement (Petrick *et al*., 1996), Fuzzy logic (Gavrielides *et al*., 2000), Morphological operations (Li *et al*., 2001) and Dynamic programming based techniques (Domínguez *et al*., 2007).

The presence of the pectoral muscle in mammogram images effects results of intensity based image processing methods and can bias procedures in detection of malignant and benign abnormalities (Kwok *et al*., 2004), (Raba *et al*., 2005), (Nicolaou *et al*., 2008), (Xu *et al*., 2007), (Ferrari *et al*., 2004), (Mirzaalian *et al*., 2007), (Bajger *et al*., 2005). Thus, during mammogram analysis the pectoral muscle should be suppressed. Figure 3.2 indicates the segmentation of the pectoral muscle (white) from the breast region (grey). In Section 3.4.3 a review on mammogram segmentation techniques widely used in this field of research is provided.

3.3.3 Feature Extraction

During feature extraction, heuristics are computed (calculated) from the characteristics of the segmented ROIs. The number of features (heuristics) selected

for breast cancer detection reported in literature varies with the approach employed. Features in different image domains such as: morphological, spatial, texture can be extracted from digital mammograms (Thangavel *et al.*, 2005a). During feature extraction, the most important characteristics of the ROIs are studied and analyzed. Important characteristics of ROIs for classification of malignant and benign tumors (mass lesions and calcifications) as reported by radiologists are as follows (Veldkamp *et al.*, 2000):

(a) *Polymorphism vs. Monomorphism*: Calcifications that are malignant tend to polymorph while benign clusters are mostly characterized by monomorphous calcifications of uniform size (Lanyi, 1988).

(b) *Size and Contrast*: Benign calcifications have larger size and contrast compared to malignant calcifications.

(c) *Branching vs. Round and Oval Type*: Linear calcifications can be an indication of DCIS, since such calcifications are located in the glandular ducts. Benign calcifications are mostly round or oval in shape and are often located in the lobules.

(d) *Orientation*: Malignant calcifications often have shapes that are oriented to the nipple (Lanyi, 1988).

(e) *Number*: A MCC with very few calcifications is regarded as less suspicious. Five or more calcifications, measuring less than 1mm in a volume of one cubic centimeter, are considered to form a MCC (Popli, 2001).

(f) *Location*: About 48 percent of all cancerous processes are located in the outer upper quadrant of the breast. Lesions found in this quadrant are more suspicious (Harris *et al.*, 1996).

During the last decade a number of mammogram feature extraction techniques have emerged and were applied to digital mammography for the detection of mass lesions and MCCs. These feature extraction techniques include: texture features (Sahiner *et al*., 1998), (Mudigonda *et al*., 2000), (Hadjiiski *et al*., 2001), radial edge-gradient analysis (Huo *et al*., 1995), gray-level image structure features (Dhawan *et al*., 1996), morphological-based features (Chan *et al*., 1998), Wavelet analysis (Qian *et al*., 1999), boundary characteristics of tumors (Kobatake *et al*., 1999), Fuzzy-neural modeling (Verma & Zakos, 2001) and region registration using temporal features (Timp & Karssemeijer, 2006).

The reviewed literature indicates that texture-based features (Haralick *et al.*, 1973) have been most commonly used for the identification of mass lesions and MCCs. Texture based descriptors proposed by Haralick *et al.* (1973) and later implemented by Kramer & Aghdasi (1999), Soltanian-Zadeh *et al.* (2004), Chan *et al*. (1998) and Khuzi *et al*. (2009) have been indicated to increase the performance of machine learning algorithms (Makinacı *et al*., 2005), (Jirari *et al*., 2005) and (Mudigonda *et al.*, 2000) such as SVMs and ANNs. Thus, the approach proposed in this research uses texture-based features for the machine learning modelling. Section 3.5 discusses in detail the approach applied in texture feature analysis.

3.3.4 Feature Selection

For the purpose of pattern classification, it is desirable to use an *optimal number of features* for machine learning modeling. Since a large number of features increases the computational needs, it becomes more challenging to define accurate decision boundaries in a large dimensional space. This indicates that an optimal subset of features needs to be selected for the purpose of machine learning.

Feature selection is an important part of any machine learning task. The success of a classification scheme mainly depends on the features selected and the information they provide for their role in the model. Some of the features extracted from the ROIs in the mammographic images are not significant when observed alone, but in combination with other features they can be significant for classification. In general, the reason for performing feature selection is three-fold namely: (a) improving the classification performance of the system, (b) providing faster and more cost effective classification and (c) providing a better understanding of the processes that generated the data (Guyon & Eliseeff, 2003).

During the last decade a number of mammogram feature selection techniques have emerged and have been applied to digital mammography for the detection of mass lesions and microcalcifications. These feature selection techniques include: Genetic Algorithm (GA) (Sahiner *et al*., 1996), (Zheng *et al*., 1999), Linear Discriminant Analysis (LDA) and linear regression analysis (Huo *et al*., 2000), Genetic programming (Nandi *et al*., 2006) and Neural-genetic modeling (Verma & Zhang, 2007).

There are many notable benefits of heuristic and feature selection. Firstly, feature selection facilitates data visualization and understanding and reduces the storage requirements by reducing the training time and improving classification performance. The discrimination power of the features can be analyzed through this process. The goal is to eliminate a feature if it gives little or no additional information, beyond that subsumed by the remaining features (Koller & Sahami, 1996). In addition, features with high correlation can be eliminated during this process in order to reduce the overall processing time without affecting the accuracy of the classifier. Only a few features may be useful or 'optimal' while most may contain irrelevant or redundant information that may result in the degradation of the classifier's performance. Irrelevant and correlated feature attributes are detrimental because they might contribute noise and can interact counter-productively to a classifier induction algorithm (Hsu *et al.*, 2002).

There are several feature selection techniques that have been well researched and published. All these methods determine the relevancy of the generated features towards the classification task. There are five main types of evaluation functions (Dash & Liu, 1997):

1. Distance (Euclidean distance measure)
2. Information (Entropy, information gain etc.)
3. Dependency (Correlation coefficient)
4. Consistency (Minimum features bias)
5. Classifier Error Rate (based on the classification algorithm)

The limitation of all the methods listed above is that they may lead to the selection of a redundant subset of features. Chen and Lin (2006) indicated that variables that are independently and identically distributed are not truly redundant. Noise reduction and better class separation can be obtained by adding variables that are presumably redundant. Chen and Lin (2006) reportedly indicated that a variable which is completely useless by itself, can provide a significant improvement in the performance when it is considered with other variables. In other words, two variables that are useless by themselves can be useful together. Thus, selecting subsets of variables together can provide good prediction results, as opposed to ranking the variables according to their individual predictive power. Thus, the approach proposed in this research applies the technique developed by Chen and Lin (2006), which uses F-score and the Random Forest (RF) with SVMs for the purpose of feature selection. This approach is discussed in detail in Section 4.4 of this thesis.

3.3.5 Classification

During the *classification* stage, the patterns of the ROIs (abnormalities in breast tissue) are classified as either benign or malignant on the basis of the optimum subset texture features selected for machine learning modeling. A classifier trained on known abnormalities (mass lesions and microcalcifications/MCCs) combines the selected features and uses confidence measures to indicate that a ROI is either malignant or benign (Duda *et al.,* 2001), (Bishop, 1995), (Ripley, 1996), (Fukunaga, 1990), (Vapnik, 1995), (Vapnik 1998).

During the training stage of learning machines, techniques to evaluate the learning and memorization performance of a classification engine (classifier) are identified,

which include: Cross-validation (CV), leave-one-out scheme, as discussed in Section 4.1.3.

Several automated classification techniques have been investigated for the detection of mass lesions and MCCs in mammograms during the last decade. These classification techniques include: Support Vector Machines (SVMs) (Wei *et al*., 2005), Artificial Neural Networks (ANNs) (Wu *et al*., 1992), (Chan *et al*., 1995a), (Jiang *et al*., 1996), (Sahiner *et al*., 1996), (Chan *et al*., 1997), (Huo *et al*., 1998), (Papadopoulos *et al*., 2002), (Zhang *et al*., 2005), LDA (Chan *et al*., 1995b), (Zhang *et al*., 2005), Convolutional Neural Networks (CNNs) (Sahiner *et al*., 1996) and the k-Nearest neighbor (Veldkamp *et al*., 2000). Other techniques include a statistical method based on the use of statistical models and the general framework of Bayesian image analysis was developed by Karssemeijer (1993).

The authors in Yoshida *et al.* (1996) used the decimated wavelet transform and a supervised machine learning technique for the detection of mass lesions and MCCs. Suzuki *et al*. (2005) proposed a method for distinction between benign and malignant tumors using ANNs. The authors in Cheng *et al*. (1998) proposed a fuzzy logic based approach for classification between different breast abnormalities. The authors in Hadjiiski *et al*. (1999) applied a hybrid combination of the Adaptive Resonance Theory (ART) and LDA for classification between malignant and benign abnormalities.

Recent studies have shown the superiority of SVMs over other supervised machine learning techniques such as the ANN, suggesting that the SVM is a promising technique for classification of noisy data. The authors in El-Naqa *et al*. (2002) used

SVMs to detect mass lesions and microcalcifications based on finite image windows. Their approach relies on the capability of SVM to automatically learn relevant features for optimal detection. In their work, a sensitivity of as high as 98 percent was achieved. Since then, SVMs have been proven to be useful for the classification of masses and MCCs in digital mammography applications, as indicated by the research groups, namely, Martins *et al.* (2009), Dehghan *et al.* (2008), Papadopoulosa *et al.* (2005), Gorgel *et al.* (2009), Wei *et al.* (2005) and Manzano-Lizcano *et al.*, (2004). Thus, the approach proposed in this research applies SVM for the purpose of pattern classification between malignant and benign texture features. The theoretical background of SVM is discussed in detail in Section 4.2 of this thesis.

3.3.6 Performance Evaluation

In order to evaluate the performance of pattern classification systems (developed using supervised machine learning techniques such as ANNs and SVMs), the binary classification performance has to be measured.

Table 3.1: Relation between, TP, TN, FP and FN — Confusion matrix

Confusion Matrix	Positive (p^a)	Negative (n^a)
Positive (p^P)	**True Positive (TP)**	**False Positive (FP)**
Negative (n^P)	**False Negative (FN)**	**True Negative (TN)**

The performance of a binary classifier cannot be described by a single value and is usually quantified by its *accuracy* during the test phase, i.e., the fraction of misclassified points on the test set. The performance of a binary classifier can be

best described in terms of its *sensitivity* and *specificity*, quantifying its performance to false positive (FP) and false negative (FN) instances (Veropoulos, 2001).

In a Receiver Operator Characteristics (ROC) curve, the *sensitivity*, which in this research is the portion of malignant tumors that are correctly classified by the learning machine, is plotted against 1-*specificity*, the share of benign tumors that are falsely classified by the learning machine, for different cut-off values. The ROC analysis generally is used to determine an optimum cut-off value referred to as criterion for use in medical diagnostic tests. It is possible to achieve an optimal balance between sensitivity and specificity that is needed for a certain purposes. This can be achieved by changing the cut-off value of the system. Also, if the cost of not detecting a particular disease becomes high to society, the cut-off value can be changed to achieve a very high sensitivity, but lower specificity (Veropoulos, 2001).

Table 3.2: Binary classification performance measures

Performance Measure	Definition
True Positive (TP)	Tumor marked as malignant by a biopsy, which is also classified as malignant by the learning machine.
False Positive (FP)	Tumor marked as malignant by a biopsy, which is classified as benign by the learning machine.
True Negative (TN)	Tumor marked as benign by a biopsy, which is also classified as benign by the learning machine.
False Negative (FN)	Tumor marked as benign by a biopsy, which is classified as malignant by the learning machine.

The confusion matrix in Table 3.1 indicates the relationship between different performance indices for binary classification. For the computerized classification

of malignant and benign abnormalities in digital mammograms, the four performance indices (TP, TN, FP and FN) in Table 3.2 are calculated by comparing the predicted output from the learning machine with the real labels determined by a biopsy. Using these four performance measures, relative measurements for binary classification can be calculated. Sensitivity is defined as the ratio of tumors which are marked and classified as tumor, to all marked tumors, given by:

$$Sensitivity = \frac{\text{Positives correctly classified}}{\text{Total positives}} = \frac{\text{TP}}{\text{TP+FN}} \tag{3.1}$$

Specificity is defined as the ratio of tumors which are not marked and also not classified as tumor, to all unmarked tumors, given by:

$$Specificity = \frac{\text{Negatives correctly classified}}{\text{Total negatives}} = \frac{\text{TN}}{\text{TN+FP}} \tag{3.2}$$

The overall accuracy is the ratio between the total number of correctly classified instances and the test set size, given by:

$$Accuracy = \frac{\text{Instances correctly classified}}{\text{Total instances}} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{3.3}$$

In order to visualize ROC curves of the binary classification performance, the performance metrics *True Positive Fraction* (TPF) and *False Positive Fraction* (FPF) can be computed using:

$$TPF = (Sensitivity) = \frac{\text{Postivites correctly classified}}{\text{Total positives}} \tag{3.4}$$

$$FPF = (1\text{-}Specificity) = \frac{\text{Negatives correctly classified}}{\text{Total Negatives}} \tag{3.5}$$

In a medical diagnosis test, *sensitivity* gives the percentage of correctly classified diseased individuals and *specificity* indicates the percentage of correctly classified individuals without the disease. So, ROC curves are two-dimensional representations the relative tradeoff between the *sensitivity* (TPF) and the 1-*specificity* (FPF) of medical diagnostic test (Veropoulos, 2001), as shown in Figure 3.3.

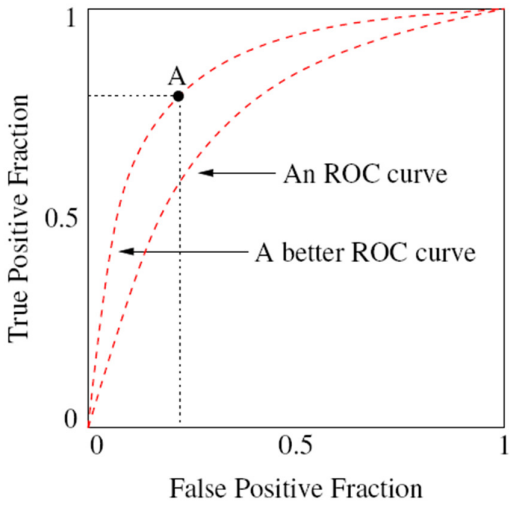
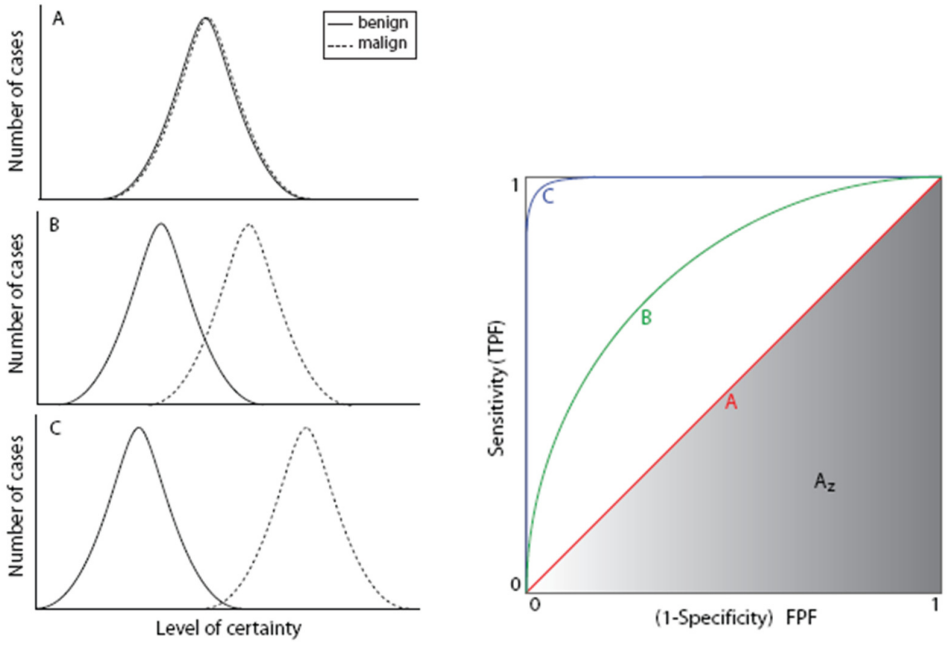


Figure 3.3: An ROC curve — FPF vs. TPF

The total *Area Under the Curve* (AUC) of a ROC curve represented by A_z, which is a quantitative measure of the binary classification performance as it reflects the testing performance of the classifier at all possible cut-off levels. The larger the AUC is within the closed interval [0.5,1], the better will be the classification performance (Veropoulos, 2001).

There are generally a finite number of points on a ROC curve in most medical diagnostic experiments, which mostly results in not finding a good approximation of the AUC. Thus, the more points there are, the better will be the estimate of the curve and the accuracy of the accuracy of the binary classifier.



(a) (b)

Figure 3.4: ROC curves (a) The likelihood of a tumor being benign or malignant (b) The ROC curves of Figure 3.4 (Veropoulos, 2001)

There are several ways to calculate the AUC of a ROC curve. The trapezoidal rule can be used to calculate the AUC, but it generally gives an underestimate of the area. Another way to get a better approximation of the AUC is by fitting the TPF and FPF data (in equations (3.4) and (3.5) respectively) into a binomial model using curve-fitting software (Veropoulos, 2001). Sample distributions of benign and malignant tumors visualized using ROC curves are shown in Figure 3.4.

The horizontal axis represents the certainty level that a tumor is malignant. When the system has difficulty in identifying if a tumor is malignant or benign, the two distributions overlap, as shown in Curve A (see Figure 3.4(a)). The AUC of Curve A is 0.5 (see Figure 3.4(b)), is the worst performance that can be obtained. Curve C in Figure 3.4(a) has the smallest overlap between the malignant and benign portions

which results in a near perfect performance with an A_z nearly equal 1.0 (see Figure 3.4(b)).

Being such a useful performance graphing method, ROC curves have been rapidly applied in several research areas, such as: medical decision-making (Veropoulos, 2001), machine learning and data mining (Spackman, 1989). In particular, ROC curves have been in signal detection theory over the past few decades to depict the tradeoff between benefits (TPs) and costs (FPs) of pattern classification systems (Egan, 1975).

3.4 Fundamentals of Digital Image Processing

The knowledge of the concepts of digital image processing presented in the following sections is required to fully comprehend the methods discussed in this thesis. In particular, emphasis is given on the image processing and segmentation techniques, which have been identified by the literature reviewed in Sections 3.3.1 and 3.3.2.

3.4.1 Representation of a Digital Image

Digital images are usually stored in a matrix form and represented as two-dimensional functions of space. Let $f(x, y)$ be this function representing an image. *Intensity function* and *Impulse function* are some of the names used to denote $f(.)$. Parameters x and y are respective row and column coordinates of a pixel in the image matrix and the value of $f(.)$ is the intensity value of a pixel. The terms *gray levels* and *intensity levels* both imply the value of a pixel in a grayscale image and are used interchangeably in this thesis.

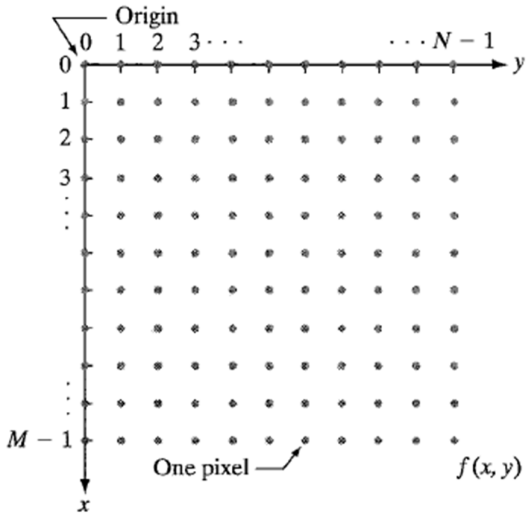


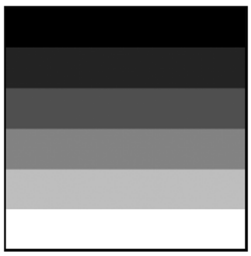
Figure 3.5: Coordinate convention used to represent digital images (Gonzalez and Woods, 2002)

The coordinate convention that is widely used to represent digital images is shown in Figure 3.5. According to the notation explained above, a digital image can be represented mathematically in a matrix form as follows:

$$f(x,y) = \begin{bmatrix} f(0,1) & f(0,2) & \dots & f(0,N) \\ f(1,1) & f(1,2) & \dots & f(1,N) \\ \vdots & \vdots & & \vdots \\ f(M,1) & f(M,2) & \dots & f(M,N) \end{bmatrix}, \tag{3.6}$$

where it is assumed that the image has M rows and N columns, or in other words, the image consists of $M \times N$ pixels, where each element of the matrix $f(x,y)$ corresponds to an image pixel. As indicated from Figure 3.5, that unlike Cartesian coordinates, x is the vertical axis and y is the horizontal axis of a digital image. All mammographic images acquired in this thesis for computer-aided modeling and experimental work consists of 1024×1024 pixels as indicated in Section 5.2.1.

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 & 4 & 4 \\ 5 & 5 & 5 & 5 & 5 & 5 \end{bmatrix}$$



(a) (b)

Figure 3.6: Grayscale image (a) A dummy matrix A (b) Image corresponding to matrix A

3.4.1.1 Range of Intensity Values

The range of intensity values for a digital image implies a closed interval $[f_{min}, f_{max}]$, where f_{min} is the smallest intensity value an image pixel can attain and f_{max} is the largest intensity value. The size of the intensity range of images during an image processing task is decided in accordance with the specifications and requirements of the task. It is a common practice to assign brighter shades to higher intensity values so that the pixel with highest value has a white shade and the pixel with lowest value has black (darker) shade. An example of an image matrix A and its corresponding gray scale image is illustrated in Figure 3.6.

When it comes to intensity range, it might seem that there are no limitations other than that intensity values can only be real values. This might be true in theory; however, due to the process involved in the acquisition of digital images, hardware considerations and other factors, the number of intensity values is typically an integer power of 2. Assuming that pixel values of an image can have L discrete gray values in the range $[0, L-1]$, and the allowed pixel values are equally spaced in this interval, then there is a positive k such that (Gonzalez & Woods, 2002):

$$L = 2^k \tag{3.7}$$

The digital mammography images acquired in this research (for the development of a framework for the computerized detection of breast cancer), have $L = 256$ gray values in the closed interval [0,255], as discussed in Section 5.2.1. If a mammogram image is of dimension $M \times N$ then the number of bits required for storage can be represented by $M \times N \times k$. A special case occurs when $k = 1$, which implies only two discrete intensity values allowed for a pixel. Such an image is also referred to as a *binary image* or a black and white image. A binary image is usually a result of the image segmentation process. In a binary image, as common practice indicates, the white shade is assigned to pixels that are *objects of interest* and the black shade is assigned to the rest of the pixels (background). For further details about image segmentation, refer to Section 3.4.3 of this chapter.

3.4.2 Histogram

A histogram is a discrete function that describes occurrence of different gray levels in an image. If the intensity range of an image $I(x, y)$ is $[0, L-1]$, then a histogram can be defined as a function $h(r_k) = n_k$, where n_k is the number of pixels that have k^{th} gray level r_k (Gonzalez & Woods, 2002). It is then possible to modify matrix A in Figure 3.6(a) in to matrix B such that:

$$B = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 5 & 5 \\ 1 & 1 & 4 & 4 & 4 & 4 \\ 5 & 5 & 5 & 5 & 5 & 5 \end{bmatrix}$$

Figure 3.7: A dummy matrix B

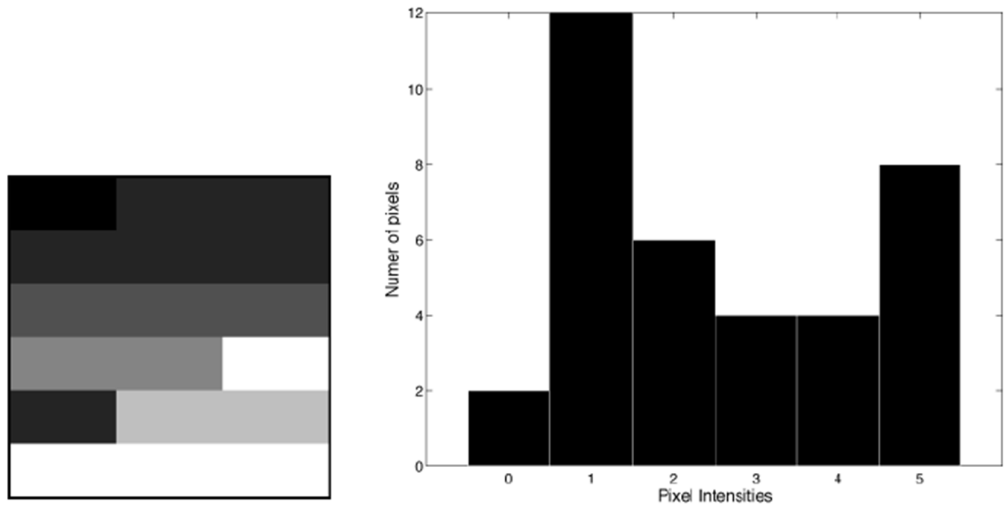
The grayscale image in Figure 3.8(a) corresponds to matrix B in Figure 3.7 where Figure 3.8(b) represents the corresponding histogram of matrix B. As observed from Figure 3.8(b), gray level 0 occurs 2 times, gray level 1 occurs 12 times, gray level 2 occurs 6 times, and so on.

3.4.2.1 Uses of Histogram

A histogram can be used to calculate and derive different properties from an image. Most of the image processing methods based on histogram analysis are statistical in nature and are related to the probability distribution of intensity values. In this thesis it has been observed that a major advantage of performing calculations on a histogram, instead of the image, is time complexity. Operations on large images will take a considerable amount of time especially if performed without optimization. On the other hand, a histogram will provide two 1D arrays, each of length L for a k-bit image as given in equation (3.7). One of these arrays contains the discrete gray levels and the other array contains the corresponding values denoting the occurrence of these gray values. Even if the images are larger than the above mentioned size, arrays obtained from the histogram are still of the same size of length L. Moreover, due to the single dimension of both these arrays, operations are relatively simple.

Even though histogram calculations may seem relatively easy and simple, there are certain drawbacks to be considered. Before being able to utilize a histogram, one has to calculate it. Calculation of an image histogram can be a time consuming operation for very large images. It is also shown in this research that methods based alone on histograms are not adequate for the primary goal of this research. Last but not the least, a histogram provides an observer with the so called global

information about an image. It is not possible to extract any local features from the histogram of an image.



(a) (b)

Figure 3.8: Grayscale image. (a) Grayscale image corresponding to matrix *B* in Figure 3.7 (b) Image histogram of grayscale image in Figure 3.8(a)

3.4.2.2 Histogram Normalization

Sometimes it is useful to normalize an image histogram by dividing each of its values by the total number of pixels in the image. This calculation is also referred to as *histogram normalization*, which creates a scaled histogram represented by $p(r_k) = n_k/n$, where n is the total number of pixels in the image. From a statistical point of view, p contains the probability distributions of different intensity values of the image.

3.4.2.3 Histogram Equalization

Histogram equalization is an interesting image enhancement tool. Figure 3.9 shows an example of histogram equalization on a grayscale image. As it is observed, Figure 3.9(a) represents an image of pollen with poor contrast and Figure 3.9(b)

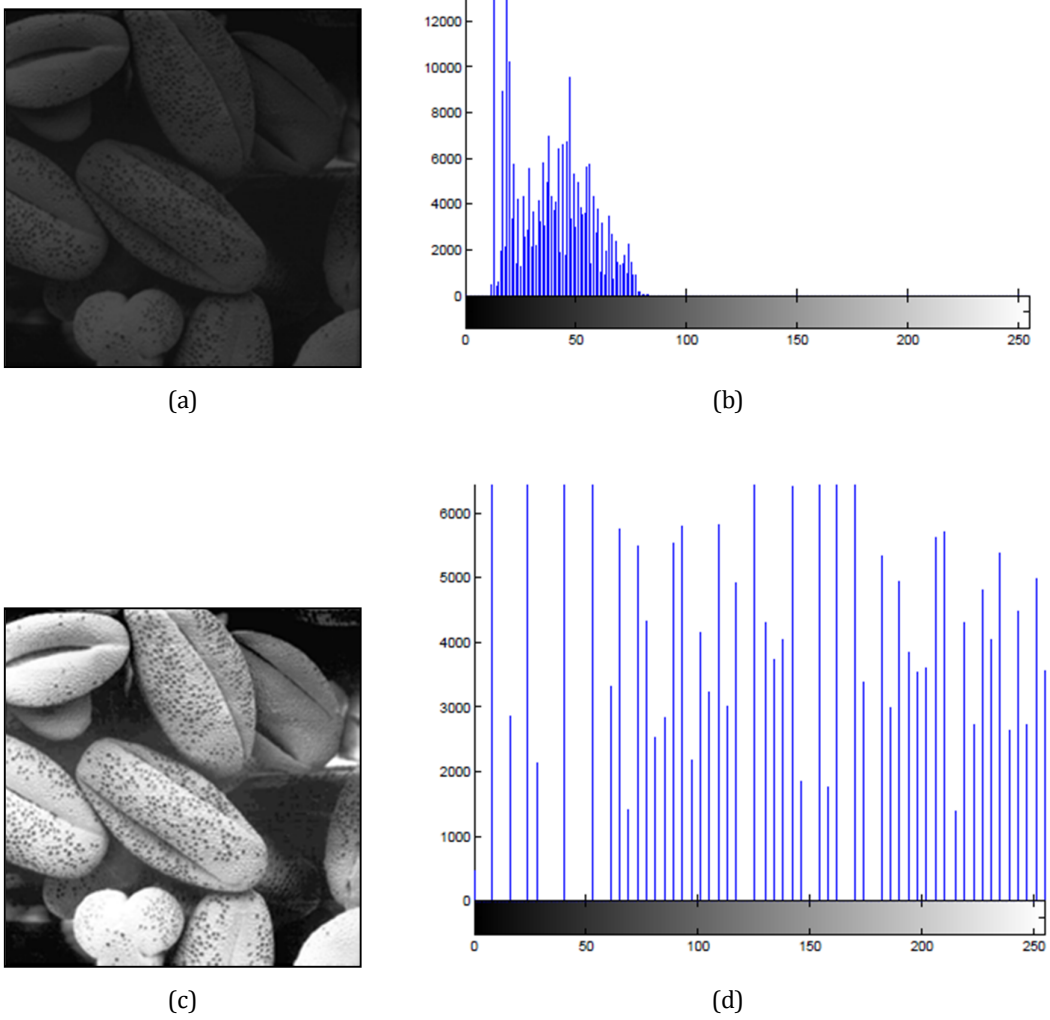
corresponds to its histogram that shows a high concentration of dark intensity pixels. If a visual analysis of the image in Figure 3.9(a) is required, it would be desirable to enhance the contrast in this image. Contrast enhancement can be achieved by applying a transformation function so that the intensity distribution becomes uniform in nature as indicated by the image histogram in Figure 3.9(d). Such transformation is called *histogram equalization*, where the corresponding grayscale image (in Figure 3.9(c)) has a brighter contrast with a balanced distribution of white and dark intensity pixels.

3.4.3 Image Segmentation

One of the most recurrent prerequisites of an image processing system is the ability to analyze images and detect regions that have specific characteristics. For instance, there is a huge demand of techniques that can enable computers to extract abnormalities and other malformations from human tissue. Such regions, in general, are called Regions of Interest (ROIs) for obvious reasons.

Image segmentation is a process that divides image pixels into smaller structural units that correspond to a ROI or neighborhood. Several segmentation algorithms are available today and the performance of these segmentation algorithms is goal specific. As explained by Rangayyan (2005), image segmentation techniques can be classified into three major categories:

1. Thresholding techniques.
2. Boundary-based methods.
3. Region-based methods.



(a) (b)

(c) (d)

Figure 3.9: Illustration effects of histogram equalization
(Gonzalez and Woods, 2002)

The following sections discuss in detail each segmentation technique and Section 3.4.3.3.1 identifies the most suitable technique.

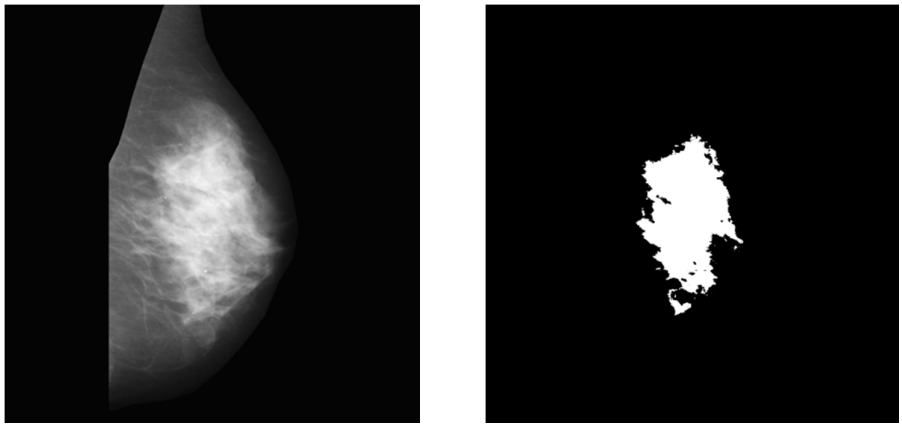
3.4.3.1 Thresholding Techniques

The main principle behind thresholding is that image pixels, falling into a predefined range of intensity values, are assigned a single intensity value, and the

remaining pixels are assigned a different intensity value. A thresholding function can be formally defined as,

$$g(x,y) = \begin{cases} i & if \quad f(x,y) \in [L_l, L_u] \\ j & if \quad f(x,y) \in [L_l, L_u] \end{cases}, \tag{3.8}$$

where $g(.)$ is the thresholded version of image $f(.)$ in equation (3.6), i and j are the two intensities used to differentiate between two groups of pixels, and L_l and L_u are the lower and upper limits of the intensity range used to define the two groups (Gonzalez & Woods, 2002). Definition of the intensity range varies depending on the images and task in hand. One can simply define a single intensity value as a threshold or define several intensity ranges. The resultant image only contains two pixel intensities, representing a binary image. For simplification purposes, in a binary image it is common to use intensity values 0 (black) and 1 (white) for i and j respectively.



(a) (b)

Figure 3.10: Segmentation using thresholding techniques. (a) Image from the MIAS database (Suckling *et al.*, 1994). Notice that the pectoral muscle has been removed to show the effects of thresholding on glandular tissue only (b) Thresholded image (Figure 3.10(a)) with a threshold value of 165

The basic principle in thresholding makes it highly suitable for segmenting ROIs that will have distinct intensities from the rest of the image. An example image taken from the Mammography Image Analysis Society (MIAS) database (Suckling *et al.*, 1994) is shown in Figure 3.10(a). The mammogram image in Figure 3.10(a) is segmented by choosing the intensity value 165 as a threshold, which as a result produces Figure 3.10(b) in which the white pixels correspond to the glandular (breast) tissue. Thus, the ROI of the image is the brighter glandular breast shown in Figure 3.10(b). The type of thresholding illustrated in Figure 3.10 is referred to as *global thresholding*, where a single threshold intensity is used to segment the grayscale image. However encouraging this result might seem, there are vital parts of the glandular tissue that have not been included in the segmentation. This is due to the fact that the type of thresholding explained here has limitations and requires additional image processing to enhance the quality of the segmentation (Rangayyan, 2005).

3.4.3.2 Boundary-based Methods

In terms of image processing, *boundary* in an image can be defined as a single continuous *edge* forming a closed path and thereby enclosing a part of the image that might be considered a ROI. A digital edge is formed by a region where transition between dark and light pixels is sharp. This phenomenon is illustrated in Figure 3.11(a) which presents a special case where the transition is rather abrupt. Boundaries made up of such an edge are easy to detect automatically, and the term *ideal edge* is used to denote them.

In boundary detection, initially an image is scanned for very sharp intensity transitions to detect edges with specific characteristics such as orientation, degree

of blurring, length, etc. Then edge-linking algorithms are applied to create enclosed boundaries. Edge-linking and boundary detection methods are explained in detail in Gonzalez and Woods (2002). Furthermore, Rangayyan (2005) and Gonzalez and Woods (2002) point out that alternative edge detection methods such as the Hough-transform based global transformation and global processing via graph theoretic techniques provide better results when it comes to creating an enclosed boundary from several disjoint sets of edge pixels.

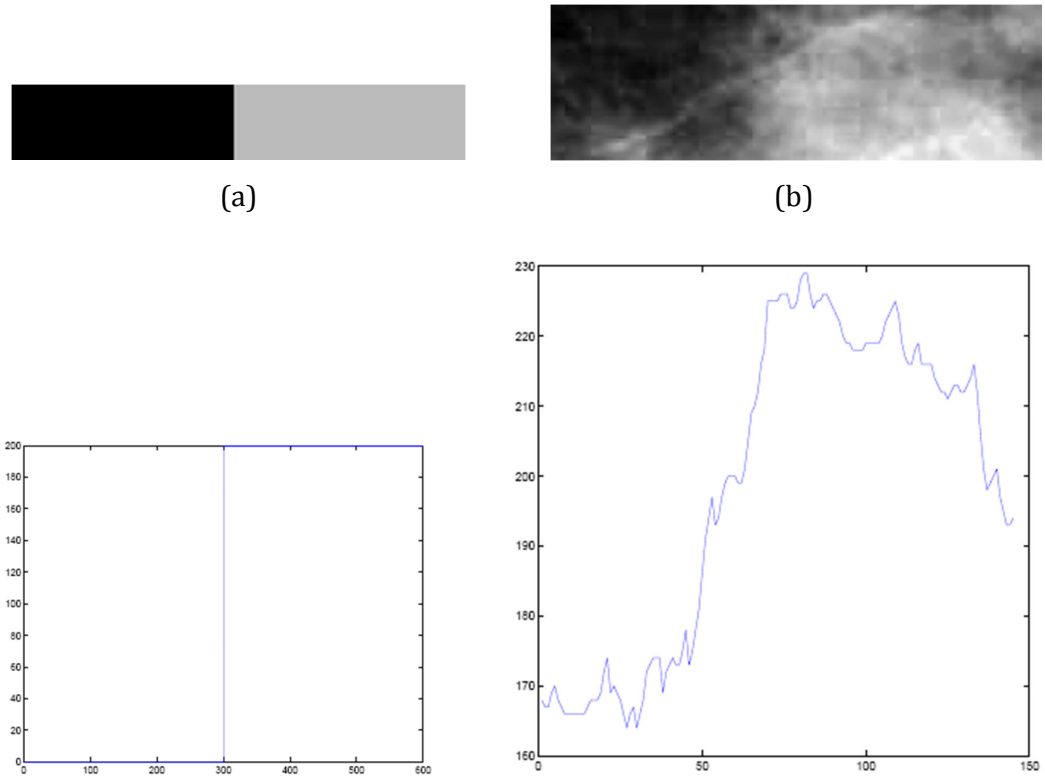


Figure 3.11: Example of an ideal edge and a blurred edge
(Gonzalez and Woods, 2002)

It is rarely the case when all ROIs are enclosed by an ideal boundary; this makes boundary detection more complex. In reality an edge will look more like in Figure 3.11(b) which is a mammogram image acquired from the MIAS database (Suckling

et al., 1994). Optics, sampling and imperfections during image acquisition are some of the reasons why edges are blurred (Gonzalez & Woods, 2002). The intensity plots of the images in Figures 3.11(c) and 3.10(d) show the transition curve of the intensity values (Gonzalez & Woods, 2002).

3.4.3.3 Region-based Methods

In most cases, regions that are to be extracted from an image have a specific texture that is different from other parts of the image. Thresholding and boundary detection might not be able to recognize specific texture because it is a local property; this is when region-based methods can be applied. In region based methods, it is assumed that pixels in a certain ROI with a specific texture share similar characteristics. A cluster of pixels sharing similar values is referred to as *neighborhood* in region-based methods. According to Rangayyan (2005) there are two types of region-based segmentation: *region splitting and merging* and *region growing*.

In region splitting and merging, an image is subdivided into smaller regions until some predefined conditions are fulfilled. For instance, a condition might be that a region should not be split if all of its pixels have the same intensity or have same fractal dimension. Then the smaller regions are merged according to some pre-specified conditions. This process might be continued until the desired result is achieved.

Region growing methods usually start with a very small group of pixels and grow a region by connecting neighborhood pixels that possess similar properties. Different properties can lead to different regions. The initial starting point of

region growing is called a *seed* pixel. It is important to choose a correct seed pixel to get the desired result. The choice of selecting a seed pixel can depend on several task specific conditions and prior assumptions. For instance, if an approximate location of the ROI is known, then the spatial centroid of this region can be used as the seed pixel. Region-based methods are often time consuming, and in some cases it is difficult to define the optimal conditions that will lead to the desired segmentation.

3.4.3.3.1 Selected Segmentation Technique

Segmentation of mammograms for the computerized detection of breast cancer can be typically performed using the three approaches discussed in Section 3.4.3. From the literature reviewed in Section 3.4.3, this section identifies the most suitable segmentation technique for the purpose of mammogram segmentation.

Thresholding techniques are based on the criterion that all pixels in an image whose pixel intensities lie within a specific range belong to a particular class (or neighborhood). Typically, thresholding methods omit all the spatial information of an image and do not deal well with noise. Similarly, boundary-based methods use the criterion that intensity values of pixels in an image change quickly at the boundary between the two regions. The basic approach used in boundary-based methods is to apply a gradient operator such as the Sobel filter (Ballard & Brown, 1982). High values of such a filter typically give candidates of region boundaries. The candidates of these boundaries need be modified so as to establish curves corresponding to the boundaries between the connected regions. Conversion of the edge pixels into boundaries of the ROIs is a challenging task. Working with regions is the complement of boundary-based approaches.

Region-based approaches rely on the criterion that all the neighboring pixels within one region have identical values (Horowitz and Pavlidis, 1974). In region-based approaches, the general task is to compare the pixel of interest with its connected neighborhood. If a criterion of homogeneity is satisfied in a region-based approach, the pixel is said to belong to the same class (or region) as one or more of its neighbors.

Segmentation techniques such as thresholding and boundary-based techniques are not able to recognize specific textures because they consider texture regions as a local property. Since the pectoral muscle in mammograms is represented as a texture property of bright contrast intensity pixels, and needs to be suppressed from the mammogram (Kwok *et al.*, 2004), (Raba *et al.*, 2005), (Nicolaou *et al.*, 2008), (Xu *et al.*, 2007), (Ferrari *et al.*, 2004), (Mirzaalian *et al.*, 2007), (Bajger *et al.*, 2005), thus, region-based methods are applied in this research to suppress un-important regions in mammogram images (Pavlidis & Liow, 1990). The theoretical concepts of an advanced region-based technique, namely, Seeded Region Growing (SRG), applied for pectoral muscle segmentation, is presented in the following section.

3.4.3.3.2 Seeded Region Growing

Region growing (Cheevasuvit *et al.*, 1986), (Pavlidis & Liow, 1990) is the most commonly used region-based method for image segmentation. Seeded Region Growing (SRG) is based on the traditional region growing criterion of the similarity of pixels within regions (Adams and Bischof, 1994). Instead of optimizing homogeneity parameters as the case with conventional region growing techniques, SRG is operated by choosing a number of pixels, known as seeds, instead of tuning

homogeneity parameters as in traditional region growing. This type of control allows unskilled and non-expert users to achieve good segmentation results on their initial attempt.

In SRG, segmentation of an image is performed with respect to a set of points, known as seeds. Consider a number of seeds grouped into n sets, say, $A_1, A_2, \dots, A_n$. In some cases individual sets will consist of single points. The choice of seeds in SRG decides that what are the features of interest and what are irrelevant or noise (Chen *et al.*, 1991). Given the seeds, the SRG tries to find a tessellation of the image into smaller regions satisfying the condition that, each connected component of a region meets exactly one of the A_i; limited to this constraint, the regions are selected to be as homogenous as possible. The description of the SRG method as applied to grayscale images is presented below (Adams & Bischof, 1994). The SRG process develops inductively from the choice of seeds selected, namely, the initial state of the sets, $A_1, A_2, \dots, A_n$. In SRG, each step of the algorithm performs addition of one pixel to any of the above sets. Then considering the state of the sets A_i after m steps, consider T be the set of all unallocated pixels, bordering at least one of the regions such that (Adams & Bischof, 1994):

$$T = \{x \notin \cup_{i=1}^{n} A_i | N(x) \cap \cup_{i=1}^{n} A_i \neq \emptyset\} \tag{3.9}$$

where $N(x)$ contains the immediate neighbours of the pixel of interest x. As an example, the immediate neighbours are those pixels which are 8-connected to the pixel of interest x. If for, $x \in T$ we have that $N(x)$ meets just one of the A_i, then we can define: (i) $i(x) \in \{1,2, \dots, n\}$ to be that index such that $N(x) \cap A_{i(x)} \neq \emptyset$, and (ii) $\delta(x)$

to be a measure of how different x is from the region it joins. The simplest definition for $\delta(x)$ is (Adams & Bischof, 1994):

$$\delta(x) = \left| g(x) - \text{mean}_{y \in A_{i(x)}}[g(y)] \right| \tag{3.10}$$

where $g(x)$ is the grey level intensity of the image pixel x. If $N(x)$ meets two or more of the A_i, $i(x)$ is taken to be a value of i such that $N(x)$ meets A_i and $\delta(x)$ is also minimized. In this circumstance, it is desirable to classify the pixel x as a boundary pixel and append it to set B, which is a set of already-found boundary pixels. We then take $z \in T$ such that (Adams & Bischof, 1994):

$$\delta(x) = \min_{x \in T}\{\delta(x)\} \tag{3.11}$$

and append z to $A_i(z)$. This process completes step $m+1$. This entire process is iteratively repeated until all pixels are allocated. In SRG, the process starts with each A_i being one of the seed sets. Thus, the definitions of $\delta(x)$ in equations (3.10) and (3.11) ensure that the segmentation (result) is into regions as homogenous as possible.

For implementing SRG using programming, a data structure termed as Sequentially Sorted List (SSL) is used. A SSL contains a linked list of objects, which contains pixel addresses that are ordered according to some feature or attribute. At the beginning of each of step of SRG, when the algorithm considers a new pixel, the pixel at the beginning of the list is taken out. When adding a pixel to the list, it is placed according to the value of the ordering attribute. In the case of SRG, the SSL stores the data of T (in equation (3.9)), which is ordered according to $\delta(x)$. The

algorithm for implementing SRG is presented in pseudo-code as follows (Adams & Bischof, 1994):

Label seed points according to their initial grouping.
Put neighbors of seed points (the initial T) in the SSL.
While the SSL is not empty:
    Remove the first point y from the SSL.
    Test the neighbors of this point:
        If all neighbors of y which are already labeled (other than with the boundary label) have the same label, then:
            Set y to this label.
            Update running mean of the corresponding region.
            Add neighbors of y which are neither already set nor already in the SSL to the SSL according to their value of δ.
        Otherwise:
            Flag y with the boundary label.

Based on the stepwise description shown above in the pseudo-code, it is observed that in executing the SRG algorithm each pixel is visited once only, although at each visit, each of the 8-connected neighboring pixels are also visited. Hence, this makes SRG a rapid algorithm.

Achieving a good segmentation performance is dependent on choosing a correct set of seeds as the homogeneity parameter (Cheevasuvit *et al.*, 1986), (Chen *et al*., 1991) and (Pavlidis & Liow, 1990). If the regions in grayscale image are noiseless, the only thing necessary for a good segmentation performance using SRG is that each pixel in a region should have a gray value which is similar to the mean of the region. In most cases, if the regions have noise present, single seeds may fall on an

outlier, which can result in a poor starting estimate of the region's mean, causing the segmentation to be incorrect. In order to prevent this from happening, it is recommended that large seed areas should be used when segmenting noisy regions in images. Furthermore, the area of each seed should be large enough so as to ensure that a stable estimate of its region's mean can be obtained (Adams & Bischof, 1994).

The choice of the *region mean* for the definition of $\delta(x)$ is assumed such that, the noise in each region contains equal variance. However, if this assumption is not true, then an appropriate choice of δ can be calculated using:

$$\delta(x) = \left| \frac{g(x) - \text{mean}_{y \in A_{i(x)}}[g(y)]}{\text{SD}_{y \in A_{i(x)}}[g(y)]} \right| \tag{3.12}$$

where SD represents the sample standard deviation of the region. In SRG, if the SD is a known function of the region's mean, then a suitable technique for variance-stabilizing should be applied to the image first.

3.4.4 Morphological Operations

Morphological operations used in digital image processing are a way of extracting image components that can be used to express details about a regions shape, its boundaries, its area and so on (Gonzalez & Woods, 2002). Two primitive and widely used morphological operations *dilation* and *erosion* are discussed in the following sections to fully comprehend the methods discussed in this thesis.

3.4.4.1 Dilation

Dilation is generally used to smooth boundaries of regions or bridge very small gaps between neighboring regions. According to Gonzalez and Woods (2002), the formal definition of dilation of a set A by another set B is denoted $A \oplus B$, and defined by:

$$A \oplus B = \left\{ z \mid \left(\hat{B}\right)_z \cap A \neq 0 \right\} \tag{3.13}$$

where $\hat{B}$ is the reflection of B. This definition means that dilation of A by B is done by reflecting B and then shifting B over A by z. Then all the displacements of B are set such that B and A overlap by at least one element, which gives the dilation. Set B is also referred to as the *dilation mask* or *structuring element* (STREL). In Figure 3.12, an example of set A and a set B are shown to illustrate the effect of dilation. The center of the mask B is marked by a small black square. In this case, the reflection $\hat{B}$ is equivalent to B. Now, if B is moved within and outside A, then dilation is given by the set of all points traversed by the center of B, until A and B are overlapped by at least one element. The resultant is shown as the shaded square that is bigger in size than A as indicated by dashed lines.

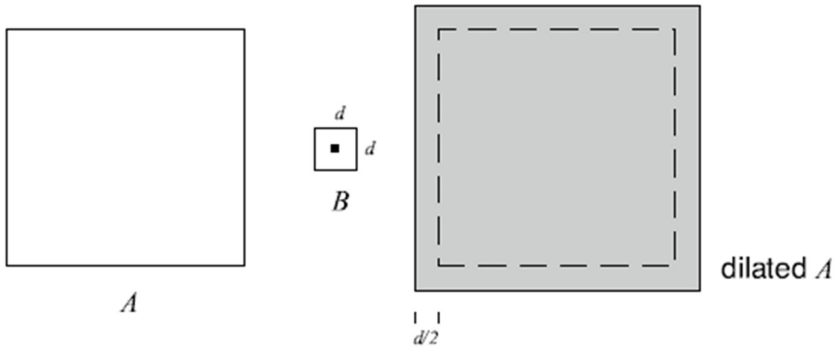


Figure 3.12: An example of dilation of set A by set B

3.4.4.2 Erosion

Erosion produces an opposite effect of dilation. Following the same notation for dilation in equation (3.13), a formal definition of erosion is given by (Gonzalez & Woods, 2002):

$$A \ominus B = \{z \mid (B)_z \subseteq A\} \tag{3.14}$$

In other words, erosion of A by B is set of all points traversed by center of B such that B is totally contained within A at all times. Figure 3.13 shows an example of set A and a set B to illustrate the effect of erosion. Erosion can be used for removing small unwanted components, such as thread like structures, from an image by using a structuring element (STREL) with an area that is bigger than the unwanted regions.

3.4.4.3 Morphological Opening and Closing

In digital image processing, the processes of morphological dilation and erosion can be combined together in different ways to make interesting changes in images. Morphological *opening* and *closing* are two such operations that are defined by specific combinations of dilation and erosion. *Morphological opening* is generally used to smooth region contours and remove thin protrusions in images. Similarly, *morphological closing* adds smoothness to image contours; however, it generally fuses two large regions separated by narrow breaks. This effect is opposite to that caused by morphological opening that breaks the narrow links between two large regions (Gonzalez & Woods, 2002).

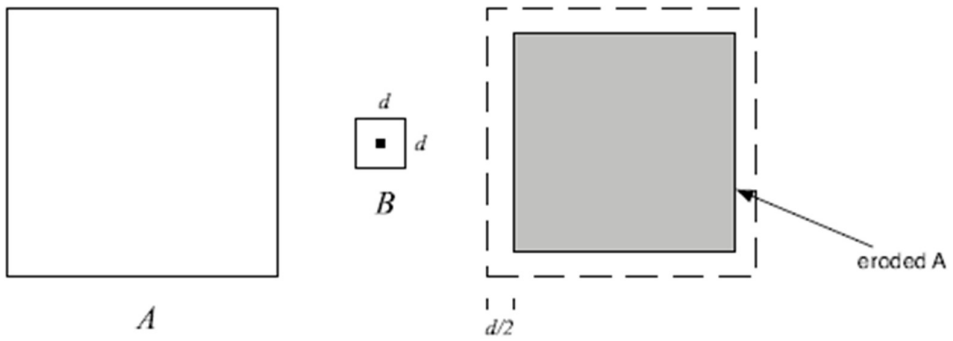


Figure 3.13: An example of erosion of set *A* by set *B*

3.5 Texture Extraction and Analysis

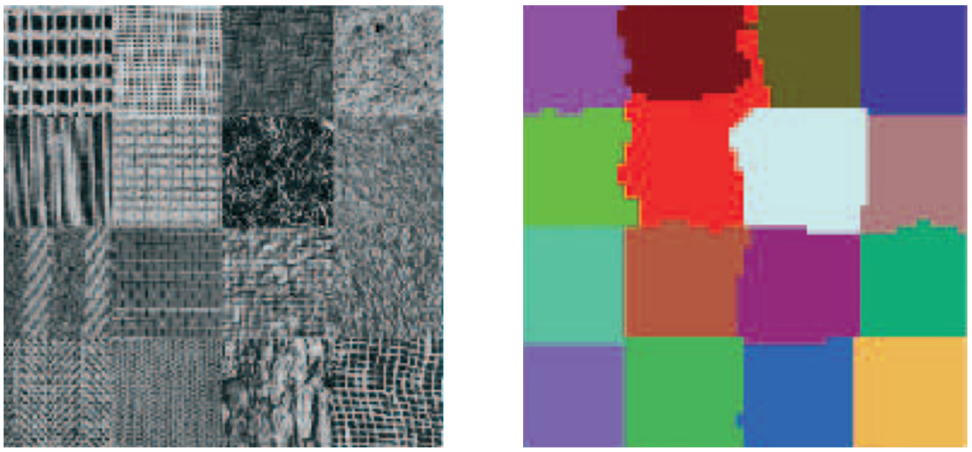
The knowledge of the concepts texture extraction and analysis presented in the following sections is required to fully comprehend the methods discussed in this thesis. In particular, emphasis is given on the texture-based feature extraction techniques and algorithms which have been identified by the literature reviewed in Section 3.3.3.

3.5.1 Introduction to Texture Analysis

Texture analysis is an important area of research in computer vision and image processing algorithms. The texture analysis domain composes of three major types of problems, which include: texture classification, texture segmentation and texture synthesis (typically used in image compression). Many definitions about texture have been formulated by different computer vision researchers. The most classical definition is as follows:

> *"Texture is defined for our purposes as an attribute of a field having no components that appear enumerable. The phase relations between the components are thus not apparent. Nor should the field contain an obvious gradient. The intent of this definition is to direct attention*

of the observer to the global properties of the display i.e., its overall "coarseness," "bumpiness," or "fineness." Physically, non-numerable (aperiodic) patterns are generated by stochastic as opposed to deterministic processes. Perceptually, however, the set of all patterns without obvious enumerable components will include many deterministic (and even periodic) textures." (Richards and Polit, 1974).



(a) (b)

Figure 3.14: Textures (a) An image consisting of sixteen different textured regions (b) Texture segmentation of (a) produced by an automatic procedure

There are three main ways of using texture features:

1. To segment an image based on texture (texture segmentation).
2. To discriminate between different (already segmented) regions or to classify them into separate categories (texture classification).
3. To produce descriptions so that textures can be reproduced (texture synthesis).

Conceptually, the main characteristic of a texture is basic pattern recurrence. The structure of a pattern can be statistical and the pattern recurrence statistically regular. Figure 3.14(a) shows an example of 16 texture categories identified as separate textures using generic labels to classify texture patterns, based on the different fill of the patterns as shown in Figure 3.14(b).

The general framework for the computerized detection of breast cancer in Figure 3.1 (Hutt, 1996) composes of a feature extraction and classification stage. Computation of texture-based features for pattern classification provides cues for classifying between patterns of malignant and benign abnormalities. As a basis for all texture related applications, texture analysis seeks to produce a general, efficient and compact quantitative description of textures so that mathematical operations can be applied to alter, compare and transform textures. The following sections discuss the most recent and successful texture-based approaches applied to digital mammography.

3.5.2 Texture Analysis Applied to Digital Mammography

This section reviews some recent publications focusing on texture analysis used in digital mammography applications and describes the contributions made. Recent studies of breast cancer aim to improve the radiologist's diagnostic performance by indicating suspicious areas. Juhl (1982) introduced the first study based on specific features sought in routine examinations as common indicators of malignancy, which supposes a big development and increment of investigation. During the last decades Chan *et al.* (1990), Davies & Dance (1990), Karssemeijer (1992) and many other research groups focused their work on microcalcification detection. Other researches such as Giger *et al.* (1990) addressed their work in

investigating methods to detect and analyze mass lesions by means of asymmetry studies. Moreover, it is important to remark that the study of Miller and Astley (1992) introduced the concept of texture analysis used to classify between breast tissues.

The increase of related works introduced by new techniques and improvements in digital mammography are mainly due to the analysis of textures, which is nowadays gaining much importance in breast cancer detection. Analysis of textures in digital mammograms is beneficial since textures have been found useful in identifying specific patterns of breast abnormalities. As digital mammography produces high resolution gray level images where textures play an important role, texture descriptors are required in order to select a set of distinguishing and sufficient texture features for the purpose of characterizing different textures in digital mammograms. Texture features contain information about the spatial distribution on the intensity pixels within defined regions in a grey scale image. Thus, the texture of a region of a digital mammogram describes the pattern of the spatial variation of grey tones in a neighborhood. Gulsrud and Loland (1996) presented an automated technique for the detection of different abnormalities in digital mammograms based on the application of multichannel-filtering for texture feature extraction.

There are several textural analysis techniques proposed in the literature for the computerized detection of breast cancer detection in digital mammography applications, however this thesis only presents the most recent and successful approaches. The three most recent texture-based approaches presented in the following sections are:

1. Grey-level Co-occurrence Matrices (GLCMs)
2. Law's Textures
3. Local Binary Patterns (LBPs)

3.5.2.1 Gray-level Co-occurrence Matrices

Gray-Level Co-occurrence Matrices (GLCMs) are one of the recent and successful texture analysis techniques commonly used for feature calculation in computer-aided applications of digital mammography.

Bovis and Singh (2000) studied detecting masses in mammograms on the basis of textural features using five GLCMs statistics extracted from four spatial orientations, horizontal, left diagonal, vertical and right diagonal corresponding to $(0°, 45°, 90° \text{ and } 135°)$ and four pixel distance $(d = 1,3,6 \text{ and } 9)$. Classification is performed using each texture feature vector and the Linear Discriminant Analysis (LDA). According to Martì *et al.* (2000), GLCMs have been frequently used in computer vision and are known to obtain satisfactory results as texture classifiers in different applications. Their approach uses mutual information with the purpose to calculate the amount of mutual information between images using histograms distributions obtained by GLCMs.

Blot and Zwiggelaar (2000a) proposed two approaches based on GLCMs for the detection and enhancement of structures in images. In the first approach, the GLCM of the local ROI is compared to a mean GLCM obtained from a number of equal size areas surrounding the local ROI. The purpose is to compare the difference between these two matrices obtaining a probability estimate of the abnormal image structures in the ROI. In The second approach by Blot and

Zwiggelaar (2000b) follows the same proposal extracting background texture in mammographic images with an improvement respect to previous works and it does not depend on any prior assumptions about the type of structures to be enhanced. Another study based on background texture extraction for classification was presented by Blot and Zwiggelaar (2001), which indicated that there is a statistical difference between the GLCM for image regions that include the image structures and regions that only contain background texture. Thus, the classification of mammographic parenchymal patterns can be improved if anatomical structures can be removed from the image.

In 2003, different approaches of GLCMs for texture-based analysis of digital mammograms were proposed. Youssry *et al*. (2003) presented a neuro-fuzzy model for fast detection of candidate circumscribed masses in mammograms, where texture features estimated using GLCMs are used to train the neuro-fuzzy model. Similarly, Martì *et al*. (2003) proposed a supervised method for the segmentation of masses in mammographic images using GLCM texture features which present a homogeneous behavior inside selected regions. Jirari (2005) proposed a computer-aided detection system for breast cancer detection using five GLCMs at different distances for each suspicious ROI and used the texture descriptors to train and test a Radial Basis Function Neural Network (RBFNN).

The authors in Costaridou *et al*. (2005) investigated how to differentiate dense breast regions containing spiculated masses from regions of normal dense tissues, by means of feature analysis on wavelet-processed mammograms. In this work Costaridou *et al*. (2005) extracted multi-resolution texture features of second order statistics from spatial GLCMs using different orientations and distances.

In more recent studies, Karahaliou *et al*. (2007) investigated texture properties of the tissues surrounding microcalcifications using a wavelet spatially adaptive enhancement method, namely the Wavelet transform. For this purpose the authors in Karahaliou *et al*. (2007) calculated thirteen textural features from four GLCMs. More recently, Lyra *et al*. (2008) investigated how to identify breast tissue quality data quantification using a computer-aided detection, where images were categorized using the BI-RADS lexicon indicated in Section 2.6.4.1. In this work, Lyra *et al*. (2008) derived texture features for each sub-region using an averaged GLCM. Karahaliou *et al*. (2008) investigated texture properties of the tissues surrounding MCCs and mass lesions in digital mammograms using gray-level texture and wavelet coefficient texture features at three decomposition levels. In modern times, Khuzi *et al.* (2009) presented a technique to identify mass lesions in digital mammograms using GLCM based texture descriptors.

3.5.2.2 Law's Texture Filter

Previous studies of Laws' textures by Miller and Astley (1992) proposed an approach to discriminate between glandular and fatty regions of breast tissue in order to automate the detection of breast asymmetries. Karssemeijer (1992, 1993) used Law's masks as a mechanism for detecting architectural distortions caused predominantly due to the ductal patterns of mammograms. Gupta and Undrill (1995) applied Law's masks to the task of delineating suspicious mass lesions and examined whether this texture-based approach indicates prospects of discrimination between stellate lesions and regular masses.

Another approach proposed by Pfisterer and Aghdasi (1998) investigated how to detect masses in digitized mammograms using textural information. In this work,

enhancement and segmentation of the images is based on texture analysis, which is performed using wavelets, steerable filters and Law's texture maps (Pfisterer & Aghdasi, 1998).

Bovis and Singh (2000) investigated a new approach for classification of mammographic images according to the breast type using Law's texture masks. In this work, Bovis and Singh (2000) extracted the total texture energy for this mask combination, for use as a feature. More work was presented by Varela *et al*. (2001) based on a digital image processing algorithm to classify mass lesions based on quantitative measures of tumor shape, contrast, and spiculation. Features based on Law's texture energy were extracted from the straightened border regions. Three Laws' filters (vertical, horizontal and symmetrical) were applied to the transformed image (Varela *et al*., 2001). In the same year, Pfisterer and Aghdasi (2001) following their previous work (Pfisterer & Aghdasi, 1998) continued their study and found better results than a specific convolutional mask, because it is not as important as the pre-enhancement used.

Karahaliou *et al*. (2006) demonstrated how texture analysis of a breast tissue surrounding microcalcifications and mass lesions shows promising results contributing to the reduction of benign biopsies. In this work, Karahaliou *et al*. (2006) computed texture features from the remaining ROI area (surrounding tissue) by using first and second order statistics algorithms, grey level run length matrices and Law's texture energy measures. Another more recent study of Karahaliou *et al*. (2007) investigated the texture properties of the tissues surrounding microcalcifications and masses using a wavelet spatially adaptive enhancement method (wavelet transform). In this work, Karahaliou *et al*. (2007)

used Law's texture energy measures for the discrimination of malignant from benign tissues were investigated using a k-Nearest Neighbor classifier.

3.5.2.3 Local Binary Patterns

Local Binary Patterns (LBPs) is a recent and promising technique for texture analysis. Oliver *et al*. (2007a) presented an approach to represent salient micro-patterns using the spatial structure of masses. The approach proposed by the authors in Oliver *et al*. (2007a) focuses on reducing the number of FPs in the field of lesion detection and distinguishes between the true recognized masses. Another work presented by Oliver *et al*. (2007b) is based on classifying mammograms using texture information. In this sense, texture descriptors are extracted from each cluster by using LBPs.

More recently, Lladó *et al*. (2007) analyzed a set of FP reduction methods in the field of mammographic mass detection using different approaches to extract lesion image features using LBPs. In modern times, Lladó *et al*. (2009) proposed the use of LBPs to characterize micro-patterns and preserve the spatial structure of masses.

3.5.3 Comparison of Texture Analysis Techniques

The literature reviewed on texture analysis techniques applied for digital mammography in Section 3.5.2 indicates that GLCMs, Law's texture filter and LPBs are the most recent texture-based approaches applied to digital mammography. Table 3.3 summarizes these textural analysis techniques with respect to the majority of work conducted and success achieved by each. As observed from Table

3.3, in modern times, GLCMs have received a notable success for computer-aided detection in digital mammography compared to Law's texture filter and LBPs.

Table 3.3: Classification of the state-of-the-art texture analysis techniques

Year	Author(s)	Feature Extraction Technique		
		GLCM	LAWS	LBP
1992	Miller and Astley (2000)	×	✓	×
	Karssemeijer (1992)	×	✓	×
1993	Karssemeijer (1993)	×	✓	×
1995	Gupta and Undrill (1995)	×	✓	×
1998	Pisterer and Aghdasi (1998)	×	✓	×
2000	Bovis and Singh (2000)	✓	×	×
	Martì *et al*. (2000)	✓	×	×
	Blot and Zwiggelaar (2000a)	✓	×	×
	Blot *et al*. (2000b)	✓	×	×
	Mudigonda *et al*. (2000)	✓	×	×
2001	Varela *et al*. (2001)	×	✓	×
	Pfisterer *et al*. (2001)	×	✓	×
	Blot and Zwiggelaar (2001)	✓	×	×
2002	Blot *et al*. (2003)	✓	×	×
	Bovis and Singh (2002)	✓	×	×
2003	Youssry *et al*. (2003)	✓	×	×
	Martì *et al*. (2003)	✓	×	×
2005	Jirari (2005)	✓	×	×
	Costaridou *et al*. (2005)	✓	×	×
	Oliver *et al*. (2005)	✓	×	×
2006	Hassanien and Slezak (2006)	✓	×	×
	Karahaliou *et al*. (2006)	✓	✓	×
2007	Oliver *et al*. (2007a)	✓	×	✓
	Karahaliou *et al*. (2007)	✓	✓	×
	Lladó *et al*. (2007)	×	×	✓
	Oliver *et al*. (2007b)	×	×	✓
2008	Lyra *et al*. (2008)	✓	×	×
	Howard *et al*. (2008)	✓	×	×
2009	Lladó *et al*. (2009)	×	×	✓
	Khuzi *et al*. (2009)	✓	×	×

Some authors such as Karahaliou *et al*. (2007) and Oliver *et al*. (2007a) have applied their approaches using two texture-based techniques, as indicated in Table 3.3. LBP is a modern and promising technique but there are still few studies where it is applied in digital mammography. Lladó *et al*. (2007, 2009) and Olivier *et al*.

(2007a, 2007b) are the most recent researchers to apply LBPs on digital mammograms. As observed from Table 3.3, previous studies on digital mammography emphasizes Law's texture filter, however in modern times GLCMs have been chosen to replace Law's texture filter (Khuzi *et al*., 2009).

3.5.3.1 Selected Texture Extraction Technique

Nowadays many techniques in computer vision deal with feature extraction in image processing. Texture features and statistical features are the most significant heuristics for the purpose of pattern recognition from digital images. Frequently used approaches for texture analysis are mainly based on statistical properties of the intensity histogram. Commonly used texture based feature extraction techniques include: autocorrelation function of textures GLCMs, Fractal texture description, Law's texture filter, LPBs etc. Many statistical feature extraction methods are present in literature, the most common include: edge frequency, primitive length (run length), mathematical morphology, Gabor transform, wavelets, etc. Other feature extraction methods are geometrical methods known as *texton*, which analyze the structure of textures by identifying basis elements. For a complete description and classification of texture analysis, refer to Gonzalez and Woods (2002).

GLCMs have received a notable success in recent years (see Table 3.3) for the computerized detection of breast cancer in digital mammography applications compared to Law's texture filter and LBPs. With the increasing number of publications on the application of GLCMs for digital mammography, it is easy to think about the importance of GLCMs compared to the other texture analysis

techniques. Thus, this research applies GLCMs for texture-based feature extraction from digital mammograms using texture descriptors discussed in (Haralick *et al.*, 1973), (Soh and Tsatsoulis, 1999) and (Clausi, 2002). The following section presents the background and theoretical concepts of GLCMs.

3.5.3.1.1 Introduction to GLCMs

The statistics of grey-level histograms give parameters for each processed region, but do not provide any information about the repeating nature of the texture. According to Beichel and Sonka (2006), the occurrence of gray-level configuration may be described by matrices of relative frequencies, called *co-occurrence matrices*. Hence, the Grey-level Co-occurrence Matrix (GLCM) is a tabulation of how often different combinations of pixel intensity values (grey levels) occur in an image. GLCMs are constructed by observing pairs of image cells at a distance d from each other and incrementing the matrix position corresponding to the grey level of both cells. This allows deriving four matrices for each given the distance: $P(0^{\circ}, d)$, $P(45^{\circ}, d)$, $P(90^{\circ}, d)$, $P(135^{\circ}, d)$. For instance, $P(0^{\circ}, d)$ is defined as follows:

$$P\left((0^{\circ}), d(a,b)\right) = \left|\left\{\begin{matrix}((k,l)(m.n)) \in D: k-m=0, |l-n|=d, \\ f(k,l)=a, f(m,n)=b\end{matrix}\right\}\right| \qquad (3.15)$$

where each P value is the number of times that: $f(x_1, y_1) = i, f(x_2, y_2) = j, |x_1 - x_2| = d$ and $y_1 = y_2$ append simultaneously in the image. Similarly, $P(45^{\circ}, d)$, $P(90^{\circ}, d)$ and $P(135^{\circ}, d)$ can be defined as follows:

$$P\left((45^{\circ}), d(a,b)\right) = \left|\begin{Bmatrix}\big((k,l)(m.n)\big) \in D: (k-m=d, |l-n|=-d)\\ \text{OR } (k-m=d, |l-n|=d),\\ f(k,l)=a, f(m,n)=b\end{Bmatrix}\right| \quad (3.16)$$

$$P\left((90^{\circ}), d(a,b)\right) = \left|\begin{Bmatrix}\big((k,l)(m.n)\big) \in D: |k-m|=d, l-n=0,\\ f(k,l)=a, f(m,n)=b\end{Bmatrix}\right| \quad (3.17)$$

$$P\left((135^{\circ}), d(a,b)\right) = \left|\begin{Bmatrix}\big((k,l)(m.n)\big) \in D: (k-m=d, l-n=d)\\ \text{OR } (k-m=-d, l-n=-d),\\ f(k,l)=a, f(m,n)=b\end{Bmatrix}\right| \quad (3.18)$$

A GLCM contains the frequency of a certain pair of pixels repetition in an image. According to equations (3.15) to (3.18), the parameters required for computing GLCMs are as follows:

- **Number of Grey Levels:** Normally, GLCMs use a grayscale image of 256 grey levels, which indicates a higher computational cost because all possible pixel pairs must be taken into account. The solution is to generate the matrix reducing the number of grayscale, and so the number of possible pixel combinations. The GLCM is always square with the same dimensionality as the number of grey-levels chosen.

- **Distance between Pixels (d):** GLCMs store the number of times that a certain pair of pixels is found in an image. Normally the pair of pixels are neighbors, but the matrix could also be computed analyzing the relation between non-consecutive pixels. Thus a distance between pixels must be previously defined.

- **Angle (θ):** Similar to the distance parameter, it is necessary to define the direction of the pair of pixels (neighbors). The most common directions are $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$ and its symmetric equivalents.

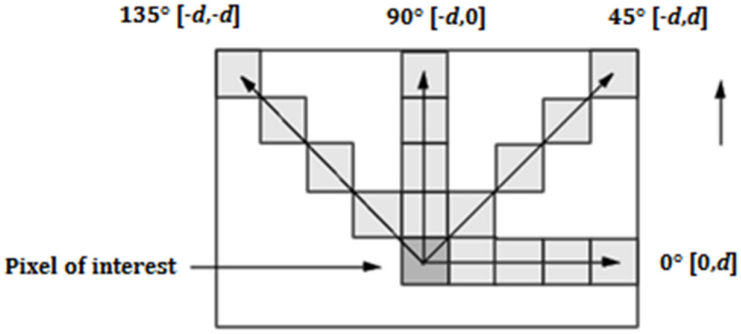


Figure 3.15: Spatial relationships of pixels defined by offsets, where d is the distance from the pixel of interest

Figure 3.15 illustrates the spatial relationships of pixels defined by the two parameters: the distance between the pixels (d) and the angle (θ). In order to illustrate the process of computation of GLCMs, Figure 3.16 shows the process used to create GLCMs. The matrix on the left in Figure 3.16 is the input image and matrix C on the right is the GLCM of the input image.

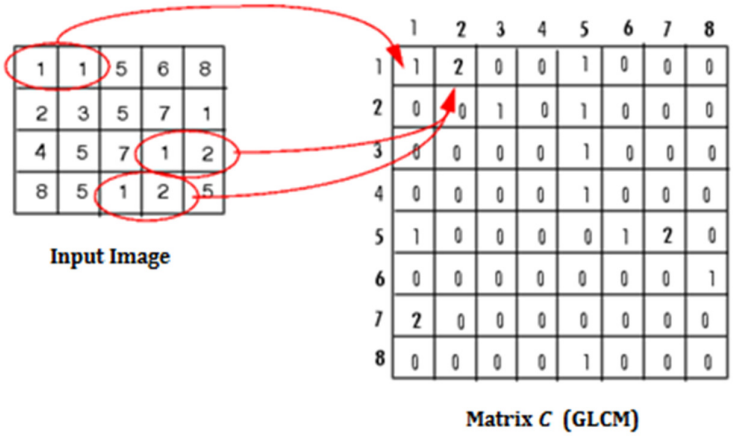


Figure 3.16: Process used to create GLCMs

As an example, consider calculating the first three values of matrix C of Figure 3.16. In matrix C, element(1,1) contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1 respectively. Similarly, element(1,2) in matrix C contains the value 2 as there are only two instances where two horizontally adjacent pixels in the GLCM have the values 1 and 2. In matrix C, element(1,3) has the value 0 as there are no instances of two horizontally adjacent pixels with the values 1 and 3 in the GLCM. The GLCM scans the rest of the input image for other pixel pairs (i, j) and the sums are recorded in the corresponding elements of the GLCM.

3.5.3.1.2 GLCM Texture Descriptors

GLCMs contain the information and properties about the spatial distribution of gray levels in grayscale images. The basis for *GLCM texture descriptors* comes from the GLCM. The GLCM is a square matrix with dimension N_g, where N_g is the number of gray levels in the grayscale image, represented in equation (3.19). Element $p(i, j)$ of matrix G (in equation (3.19)) is generated by first counting the number of times a pixel with value i is adjacent to a pixel with value j in the GLCM and then by dividing the entire matrix by the total number of comparisons made. Thus, each entry is considered to be the probability that a pixel with value i will be adjacent to a pixel with value j.

$$\mathrm{G} = \begin{bmatrix} p(1,1) & p(1,2) & \cdots & p(1,N_g) \\ p(2,1) & p(2,2) & \cdots & p(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_g,1) & p(N_g,2) & \cdots & p(N_g,N_g) \end{bmatrix} \tag{3.19}$$

Since adjacency can occur in each of the four directions (i.e., horizontal, vertical and left and right diagonals) for a two-dimensional image as shown in Figure 3.15, four matrices, each corresponding to one direction can computed as given in equations (3.15) to (3.18).

Haralick *et al.* (1979) described fourteen texture descriptors for characterizing GLCMs, which are shown in Table 3.4 (Haralick *et al.,* 1979) in equations (3.20) to (3.33). The texture descriptors proposed by Haralick (1973) form the basis for texture feature extraction using GLCMs. The term $p(i,j)$ in Table 3.4 represents the ij^{th} term of matrix C (in Figure 3.16) divided by the sum of the elements of C. In this research apart from using texture descriptors proposed by Haralick *et al.* (1979), other recent texture descriptors proposed by Soh & Tsatsoulis (1999), Clausi (2002) and the MATLAB Image Processing Toolbox are evaluated, as indicated in Section 5.4.2 of this thesis.

3.6 Summary

This chapter presented and discussed the background literature on the computer processing of digital mammograms. In Section 3.1, computer-aided detection systems were introduced with the literature review of computerized breast cancer detection techniques presented in Section 3.2. Section 3.3 identified the key techniques and algorithms used in this research to develop a framework for the computerized detection of breast cancer. Section 3.4 presented the fundamental concepts of digital image processing with emphasis on image segmentation techniques used in digital mammography applications. Lastly, Section 3.5 emphasized on the use of texture-based analysis for the purpose of feature extraction in pattern classification problems.

Table 3.4: Standard GLCM texture descriptors (Haralick, 1973)

No.	Texture Descriptor	Formula	Equation No.
1.	Angular Second Moment: Energy	$\sum_i \sum_j \{p(i,j)\}^2$	(3.20)
2.	Contrast	$\sum_{n=0}^{N_g-1} n^2 \left\{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j), \lvert i-j \rvert = n\right\}$	(3.21)
3.	Correlation (Haralick)	$\sum_i \sum_j \frac{(ij)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$ where, μ_x, μ_y, σ_x and σ_y are the means and standard deviations of p_x and p_y, the partial probability density functions.	(3.22)
4.	Sum of Squares: Variance	$\sum_i \sum_j (i-\mu)^2 \, p(i,j)$	(3.23)
5.	Inverse Difference Moment	$\sum_i \sum_j \frac{p(i,j)}{1+(i-j)^2}$	(3.24)
6.	Sum Average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$ where, x and y are the coordinates (row and column) of any entry in the co-occurrence matrix, and $p_{x+y}(i)$ is the probability of co-occurrence matrix coordinates summing to $x+y$.	(3.25)
7.	Sum Variance	$\sum_{i=2}^{2N_g} (i - S_{ent})^2 \, p_{x+y}(i)$	(3.26)
8.	Sum Entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\} = S_{ent}$	(3.27)
9.	Entropy	$-\sum_i \sum_j p(i,j) \log(p(i,j))$	(3.28)
10.	Difference Variance	$\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i)$	(3.29)
11.	Difference Entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i) \log\{p_{x-y}(i)\}$	(3.30)
12.	Information Measure of Correlation 1	$\frac{HXY - HXY1}{\max\{HX, HY\}}$ where, HX and HY are the entropies of p_x and p_y such that: $HXY = -\sum_i \sum_j p(i,j) \log(p(i,j))$ $HXY1 = -\sum_i \sum_j p(i,j) \log\{p_x(i)p_y(j)\}$ $HXY2 = -\sum_i \sum_j p_x(i)p_y(j) \log\{p_x(i)p_y(j)\}$	(3.31)
13.	Information Measure of Correlation 2	$(1 - \exp[-2(HXY2 - HXY)])^{1/2}$	(3.32)
14.	Maximum Correlation Coefficient	$\sqrt{(\text{Second largest eigenvalue of } Q)}$ where, $Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$	(3.33)

CHAPTER 4

PATTERN RECOGNITION AND FEATURE SELECTION

4.0 Overview

An overview of pattern recognition is given in this chapter, with particular emphasis on a specific machine learning technique, namely, the Support Vector Machine (SVM). SVMs will be used intensively in this research. The reason for using SVM as the main machine learning technique for this research is discussed in Sections 1.3 and 3.3.5. Section 4.1 presents some introductory notions regarding the theoretical concepts of learning machines. Section 4.2 introduces the fundamental concepts of the statistical learning theory and presents the mathematical formulation of the SVM developed by Vapnik (1998) which describes the statistical aspects of automated machine learning. Towards the end of this chapter, Section 4.3 presents the theoretical concepts of ANNs whereas Section 4.4 discusses a Recursive Feature Elimination (RFE) technique used for the selection of the optimal subset of texture features for the learning machine (SVM).

4.1 Machine Learning

Machine learning is concerned with the development of algorithms and techniques that readily allow computers to mimic intelligent behavior using empirical data, such as from sensors or databases. Data from sensors is seen as different samples that illustrate the relations between observed variables or features. A major challenge in machine learning research is to learn how to recognize complex patterns and make decisions based on the data. The only limitation in machine

learning is that, the set of all possible behaviors (patterns) given all possible inputs, is too large to be covered by the set of trained (seen) samples (Brieman, 2001). So, the learning machine must learn (training phase) from the given samples, so it can predict (classify) from new and unseen samples (testing phase). The following sections discuss in detail regarding the fundamentals concepts of machine learning and pattern recognition.

4.1.1 The Act of Learning

In humans the act of learning is namely the process of gaining knowledge or skill in something by experience. Common and apparently simple human processes such as: recognizing a landscape, understanding spoken words, reading handwritten characters or identifying an object by touching it, all belie in the act of learning (Cortes & Vapnik, 1995). In fact, the condition for a landscape to be recognized, spoken words to be understood, handwritten characters to be read and objects to be identified, is that the human brain has been previously trained in order to do that, namely it has *learned* how to do that. This is why it is necessary to admire a landscape several times before recognizing it from a slightly different view, or to hear an unknown foreign word more than once before becoming familiar with it.

From the examples discussed above, it is evident that the act of learning plays a crucial role in all the processes requiring the solution of a pattern recognition task; all the processes in which the human brain is required to take an action based on the class of the data it has acquired are referred to as learning. For instance, hearing a voice and deciding whether it is a male or a female voice, reading a handwritten character and deciding whether it is an $\mathcal{A}$ or a $\mathcal{B}$, touching an object and guessing its temperature, these are typical pattern classification problems

(Schölkopf, 1997). These processes represent the totality of the processes a human being has to deal with. Finding a solution for them has been crucial for humans to survive. For that reason, highly sophisticated neural and cognitive systems have evolved for such tasks during the last few decades. The scheme used by the human brain to address pattern recognition tasks is based on two separate phases, namely: a training phase and a testing phase. In the training phase the human brain gets experience by dealing with patterns taking from the same population, as landscapes, spoken words or handwritten characters. Then, in the test phase, it applies to patterns of the same population, which are previously unseen. In this sense, admiring a known landscape several times by trying to identify its characteristics, represents the training phase, whereas recognizing it from a slightly different view represents the test phase.

As regard to computers, the act of learning refers to artificial intelligence. In this context, machine learning is an area of artificial intelligence that deals with the development of techniques which allow machines to learn how to solve pattern recognition problems; whereas learning machines are automata which solve pattern recognition problems. In a similar way to what happens for the human brain, the solution of a pattern recognition problem initially involves the collection of a dataset of training patterns. The learning machine structure is then adapted such as to create a mapping relationship between the input and output values. The recognition performance of the trained learning machine is then evaluated on a data of test patterns, namely patterns which were not part of the training dataset, but were taken from the same population (Tarassenko, 1998).

The success of machine learning since the 1960s up to modern times is two-fold. Firstly, it is evident that implementing learning processes by using machines is fundamental in order to automatically address pattern recognition problems, which due to their complexity are challenging for humans to solve. For example, pattern classification and recognition problems such as speech recognition, fingerprint identification, handwriting recognition, face recognition, DNA sequence identification, video surveillance, and much more can be easily addressed by means of learning machines. Secondly, by trying to give answers and explanations to the numerous questions and doubts arising from the implementation of such automatic learning systems, a deeper understanding of the processes governing human learning is gained (Brieman, 2001).

As already discussed, all those problems requiring human or artificial intelligence to take actions based on the data acquired are formally defined as pattern recognition problems. This family of problems can be further divided into families of sub-problems. The most common and important of these are *pattern classification*, *regression* and *time-series prediction* problems (Tarassenko, 1998).

Pattern classification problems are those in which the learner is required to learn how to separate the input patterns into two or more classes. A typical pattern classification problem can require, for example, a human brain or a learning machine to separate between two classes of patterns, such as malignant and benign abnormalities taken from digital mammography datasets, as shown in Figures 5.23 and 5.24 respectively. When the problem does not require associating the class of membership to input patterns, but rather to associate a continuous value, a regression problem is faced (Schölkopf, 1997). A typical regression

problem could require a human brain or machine learning algorithm to predict prognosis as a regression of the factors associated with breast cancer using clinical datasets. Lastly, time-series prediction problems, in which a learning machine is trained to predict the $(n+1)^{th}$ sample in a time-series from the previous n samples is a special case of regression problems, which assumes that the underlying data generator is stationary, namely its statistical properties are time-independent (Schölkopf, 1997). In this research, attention will be concentrated on pattern classification, which is the most common type of pattern recognition problem.

4.1.2 Learning Pattern Classification

Specific details about pattern classification, namely the way in which learning machines address pattern classification tasks is presented in the following sections. In particular, two important aspects of learning machines will be discussed. First, how they learn directly from data, without using any priori assumption on the classification problem they are facing. Secondly, how supervised and unsupervised learning paradigms are implemented in order to practically solve pattern classification problems.

4.1.2.1 Learning from Data

One of the most important characteristics of learning machines is that they are not programmed by using some prior knowledge on the probability structure of the dataset considered; in fact they are trained using repeatedly large numbers of samples for the problem under consideration. In a sense, they learn directly from the data how to separate the different existing classes. This approach determines some important peculiarities of learning machines. First, they are particularly

suited for complex classification problems, as the whole solution is difficult to specify a priori. Secondly, after being trained they are able to classify data previously not encountered. This is often referred to as the *generalization* ability of learning machines. Finally, since they learn directly from data, so the effective classification solution can be constructed far more quickly than using traditional approaches.

For this reason, an approach purely based on learning from data is regarded as the most appropriate solution. For example, as described in Duda *et al.* (2001), the Bayesian decision theory is considered a fundamental approach for solving pattern classification problems. This theory assumes that the decision problem is given in probabilistic terms and all the probability values are known. The Bayesian theory quantifies the tradeoffs between different classification decisions using probability estimates and the costs that accompany these decisions. Unfortunately, for the most part of applications, the probabilistic structure of the problem is unknown.

Since, there is limited knowledge about the situation, the training data can be used to classify the patterns. The approach is then to determine a way to use this information in order to design the classifier. One approach is to use the training patterns for estimating probabilities and densities, and use the resulting estimates as they are the true values (Schölkopf, 1997). For this reason, with regards to the classification of malignant and benign patterns in this research, it would be more appropriate and effective to address the classification task using a modeling approach rather than an approach purely based on learning from data.

Pattern classification in machine learning applications can be implemented using two major schemes, namely: *supervised learning* and *unsupervised learning*. In a supervised learning scheme, the input patterns used to train the learning machine are labeled, in other words they are patterns whose class membership is known. However, in an unsupervised learning scheme, the input patterns are unlabeled, namely their class membership is unknown Duda *et al.* (2001).

The secondary contribution of this research outlined in Section 1.3 of this thesis, is to demonstrate that SVMs can effectively solve pattern classification problems. The SVM being a supervised learning scheme is discussed in detail in Section 4.2 of this thesis. The following section provides the fundamental concepts and mathematical details of the supervised learning scheme, which is required to fully comprehend the methods discussed in this thesis.

4.1.2.2 Supervised Learning

Supervised learning is defined as the machine learning task of inferring a function from supervised training data. As the training data consists of a set of training samples, each training sample represents a pair of an input features (typically a vector) and a desired output value or label (referred as the class membership). Similarly, each testing sample is a pair of input features with no desired output value (class membership) (Cortes and Vapnik, 1995).

In pattern classification problems, a supervised learning algorithm analyzes the training data and produces an inferred function, also known as a *classifier*. During the testing phase, the inferred function is used to classify the correct class membership for any valid pair of input features. Thus, the inferred function

predicts the class membership of the testing samples. This requires the learning machine to generalize using the training data to classify unseen samples in an appropriate manner (Vapnik, 1995). In order to implement the supervised learning scheme for a binary (two-class) classification task, a labeled dataset of training patterns must be provided. To serve this purpose, the training patterns (Vapnik, 1998):

$$(x_1, \dots, x_l) \qquad \text{with} \qquad x_i \in \mathbb{R}^n \quad \forall i = 1, \dots, l \tag{4.1}$$

are required, as well as the associated labels indicating their class membership:

$$(y_1, \dots, y_l) \qquad \text{with} \qquad y_i = \pm 1 \quad \forall i = 1, \dots, l \tag{4.2}$$

Then each training pattern i is represented by a vector x_i of n input features, namely the individual measurable heuristic properties of the malignant and benign patterns. For the case of this research, the heuristic properties of the patterns refer to the classification features, namely texture features, as indicated in the proposed framework in Figure 1.4. The classification features of each sample are associated to a class membership of the label y_i, which takes values $+1$ or -1 for a binary classification problem. For example, in a pattern classification problem in which malignant and benign samples (see Figures 5.23 and 5.24 respectively) are required to be separated classification features such as the: pixel intensity values of each or some specific measurements of each image, such as: luminosity, gradient etc. can be used. As regards to the class memberships (or labels) in this research, all patterns belonging to the category (class) of $malignant$ patterns are associated to the label $+1$ and those belonging to the category of $benign$ patterns are associated to the label -1, as shown in Figure 4.1 (Cortes and Vapnik, 1995).

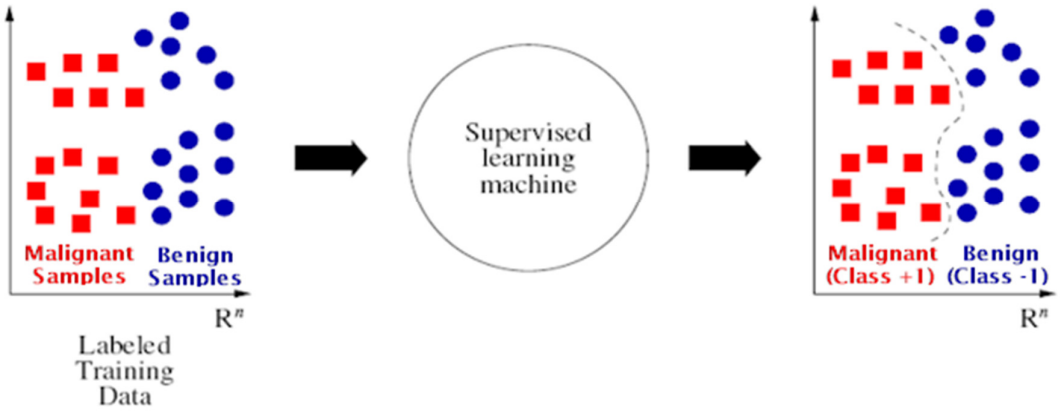


Figure 4.1: Supervised learning scheme

During the training phase of the supervised learning scheme, the learning machine adjusts its internal parameters by being shown the input features x_i of the patterns taken from *malignant* class and those patterns taken from the *benign* class. Once the training phase terminates, the learning machine is supposed to have learned how to recognize features belonging to both classes (Cortes and Vapnik, 1995). In particular, it is supposed to have learned how to correctly separate the two different classes in the n-dimensional feature space, as shown in Figure 4.1. At the end, the generalization performance of the classifier is tested on a new set of unseen data (testing data), which is not part of the training set.

4.1.2.3 Issues in Pattern Classification

In order to improve the performance of learning machines, it is most commonly desirable to submit the data to some preprocessing techniques whose aim is to enhance the data and reduce noises, so that the learning machine can provide optimum classification results. Even though preprocessing techniques are not a part of the classification task, they play a fundamental role in it. For that reason, they are usually considered as sub-issues of pattern classification problems (Duda *et al.*, 2001), (Tarassenko, 1998).

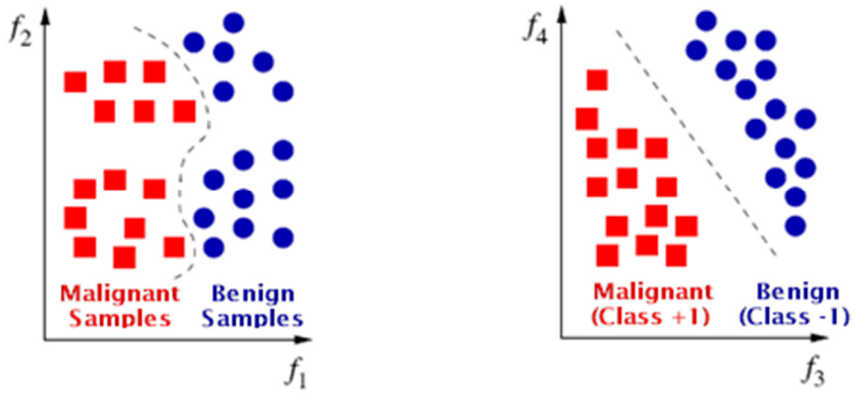


Figure 4.2: Feature extraction. The classification problem is more easily separable using the pair of features $f_3$and f_4 (right) than using f_1 and f_2 (left)

When dealing with pattern classification problems, the first important preprocessing step must be taken in the direction of choosing the most appropriate features. This literally means selecting the measurable properties of the phenomena under consideration, which can be the most discriminant features for the specific pattern classification problem faced. For example, as indicated in Figure 4.2, it could happen that the choice of a pair of features, say, $f_3$and f_4, makes the classification task more easily separable than using a different pair of features, namely f_1 and f_2. It is evident that extracting the most discriminant feature is a problem, which requires a priori knowledge of the data. Furthermore, the conceptual boundary between feature computation and classification can be considered as arbitrary. As an ideal feature extraction would yield a representation that makes classification trivial, similarly, a powerful classifier would not need the assistance of a feature extractor (Duda *et al.*, 2001). Section 4.4 presents and discusses feature selection using SVMs in detail.

The second preprocessing step consists in removing the presence of noise in the data. The definition of noise is very general. Any property of the sensed patterns

due to randomness in the world or in the sensors can be considered as noise. All non-trivial decision and pattern classification problems involve noise in some form, since all real-world data is transduced from sensors into digital data. It is thus unavoidable that all sensory datasets suffer from the specific noise of the experimental set up. Typical examples are visual noise in video cameras, background noise in audio registrations and so on.

4.1.3 Validation Techniques

The introduction of validation techniques is motivated by the willingness of finding a solution to two fundamental problems in pattern classification. The first problem being the selection of the learning machine's model and the second problem being the validation of the learning machine's classification performance (Kohavi, 1995). Almost invariably, all learning machines have one or more free parameters which can be tuned in order to adapt them to each specific classification problem. For example, in Artificial Neural Networks (ANNs) the free parameters are represented by the number of layers in the network and by the weights linking each input pattern to each perceptron. When a pattern classification problem is addressed by learning machines, the typical approach consists in choosing a specific configuration of the free parameters, namely choosing a specific model for the learning machine and estimating its classification performance. Classification performance is usually estimated by the so-called *true error rate*, literally the learning machine's error rate on the entire population under examination. The configuration of free parameters for which the true error rate is minimum, corresponds to the optimal learning machine's model for that particular problem (Tarassenko, 1998).

It is evident that in an ideal and unrealistic situation in which the dataset is comprised of an unlimited number of patterns, the straightforward solution of the problem will be at first to choose the learning machine's model that provides the lowest error rate on the entire dataset, and this error rate can be considered as the true error rate. Obviously, in real-world applications, only finite datasets are available and typically they are smaller than what would be desirable.

In this case, a crude approach would be to use the entire dataset to train the learning machine in order to select the model and to estimate the error rate. However, this crude approach suffers from two fundamental drawbacks. First, the final model will normally *overfit* the training data. This means that the learning machine results in excessive optimization on the training data, thus, it loses in generalization performance and gives very poor interpolation on a different dataset. Secondly, the error rate estimate is overly optimistic, typically lower than the true error rate. It is in fact not uncommon to have 100 percent correct classification on the training data (Duda *et al.*, 2001). In order to overcome these drawbacks, some sophisticated validation techniques for supervised learning schemes have been introduced, which are presented in the following sections.

4.1.3.1 Holdout Method

An interesting approach consists of splitting the dataset into two disjoint subsets, thus applying the *holdout* validation technique. The holdout method, generally referred to as the *test sample estimate*, partitions the data into two mutually exclusive subsets, known as the training set and the test set, in analogy to what has been discussed in the previous sections. It is common to designate two-thirds (77.5 percent samples) of the dataset as the training set and the remaining one-third

(33.5 percent samples) as the test set, as indicated in Figure 4.3 (Kohavi, 1995). The training set is used to train the learning machine and the trained learning machine is then tested on the test set.

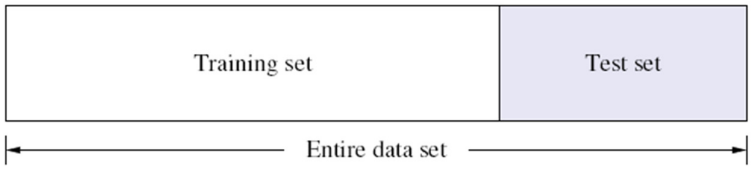


Figure 4.3: Holdout method

This method suffers from two important drawbacks. Firstly, assuming that the learning machine's classification performance increases as more patterns are seen, the holdout approach is a pessimistic estimation routine because only a portion of the data is given during the training phase. Secondly, since it is a single train and test experiment, its estimate of the error rate can be inaccurate and misleading if it happens to get an unfortunate split, i.e., if the test set is composed of the most difficult patterns from the entire dataset.

4.1.3.2 Cross-validation

The k-fold Cross-Validation (CV) is a technique in which the data set $\mathcal{D}$ is randomly split into k mutually exclusive folds $F_1, F_2, \dots, F_k$ each having an equal size. In a specific CV case, referred to as *stratified cross-validation*, the folds are stratified so that they contain the same proportion of the samples as in the original dataset. Here the learning machine is trained and tested k times, namely for each time $t \in \{1,2, \dots, k\}$ it is trained on $\mathcal{D} \backslash \mathcal{D}_t$ and tested on $\mathcal{D}_t$, as shown in Figure 4.4 (Kohavi, 1995).

The major advantage of the CV technique with respect to the holdout method is that all the patterns in the dataset are used for training and testing. At the same time, the true error is estimated as the *average error* on the test patterns, thus preventing the problems arising from unfortunate splits. The average error rate for k-fold CV can be calculated using the following expression (Kohavi, 1995):

$$e_{avg} = \frac{1}{k}\sum_{i=1}^{k} e_i \tag{4.3}$$

where k represents the number of folds and e_i represents the true error rate for each of the k-folds.

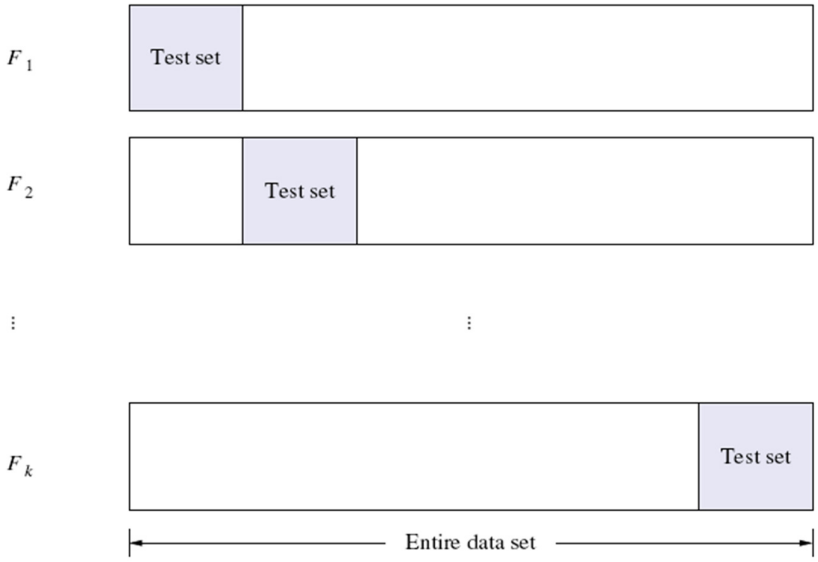


Figure 4.4: k-fold cross-validation method

Some interesting considerations concerning the choice of the correct number of folds k are drawn in Kohavi (1995). First, if the number of folds k is large, the bias of the true error estimate is generally small, thus the estimator may be considered very accurate. Unfortunately, due to the large number of iterations, the variance of the true rate estimator as well as the computational time is expected to be large.

Secondly, if the number of folds k is reduced, the bias of the true error estimate is generally large, thus the estimator may be considered conservative or higher than the true error rate. In such cases, due to the reduced number of iterations, the variance of the true error rate estimator as well as the computational time is typically small. In practice, the choice of the number of CV folds strongly depends on the size of the dataset. For large datasets, even a 3-fold CV could be quite accurate. For sparse datasets, it may be necessary to partition the dataset in a large number of folds to train on as many patterns as possible. A common choice for k-fold cross validation is typically, $k = 10$.

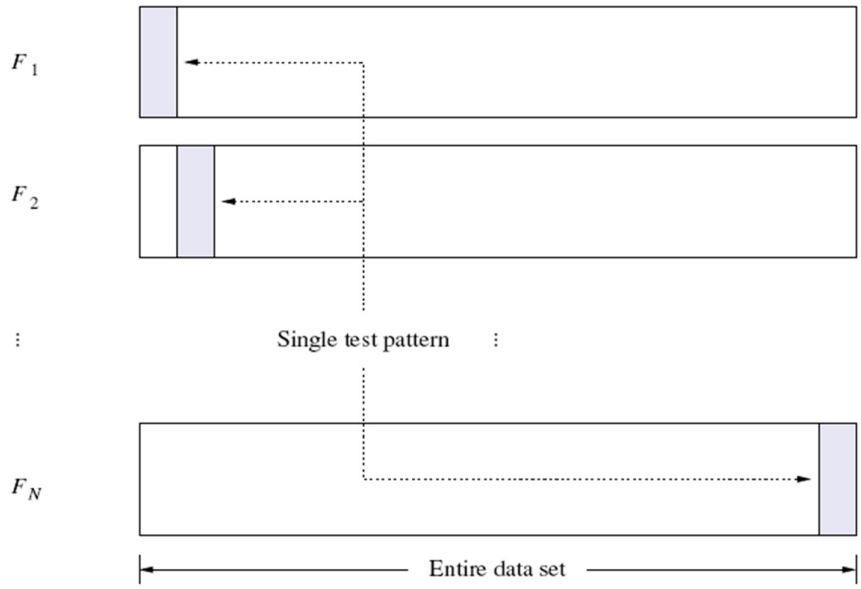


Figure 4.5: Leave-one-out method

4.1.3.3 Leave-One-Out Method

For very sparse datasets, the *leave-one-out* validation method is simple to implement. This approach is analogous to that of CV. The only difference is that, if N is the number of patterns in the dataset, the learning machine is trained N times. In particular, for each time, $N - 1$ patterns are used for training and the remaining

for testing, as shown in Figure 4.5. Similar to CV, the true error rate in *leave-one-out* validation is estimated as the average error rate on the test patterns Kohavi (1995).

4.2 Support Vector Machine

The Support Vector Machine (SVM) is one of the advanced techniques amongst the many learning algorithms deeply inspired by the statistical learning theory, which appeared in the machine learning community in the last decade. Its original formulation is quite recent and is mainly due to Vapnik & Chervonenkis (1974), Boser *et al.* (1992), Guyon *et al.* (1993), Cortes & Vapnik (1995), Vapnik (1995, 1998).

During the 1990s many machine learning algorithms arose contradicting the biological paradigm, since most were inspired by the minimization of theoretical bounds on the error rate. The SVM is not an exception to that. In fact, for a given learning task and with a finite amount of training patterns, the SVM is a learning machine which achieves its best generalization performance by finding the right balance between the accuracy obtained on that particular training set and the complexity of the machine, namely its ability in learning any training set without errors (Burges, 1988).

The following sections present introductory notions on the statistical learning theory and the mathematical concepts of SVMs, in order to demonstrate that finding the right balance between accuracy and capacity is equivalent to finding the minimum of a theoretical bound on the error rate.

4.2.1 Statistical Learning Theory

Suppose that, l training patterns are represented by (Vapnik, 1998):

$$x = (x_1, \ldots, x_l) \quad \text{with} \quad x_i \mathbb{R}^n \quad \forall i = 1, \ldots, l \tag{4.4}$$

together with the associated class labels representing the attended class membership:

$$y = (y_1, \ldots, y_l) \quad \text{with} \quad y_i = \pm 1 \quad \forall i = 1, \ldots, l \tag{4.5}$$

Assume that patterns are generated independently and distributed identically according to an unknown probability distribution $P(x, y)$. As already known, learning machines address the task of pattern classification by finding a rule which assigns each input pattern to a class membership. In particular, during the training phase, a mapping $f: \mathbb{R}^n \rightarrow \{\pm 1\}$ is created between input patterns and labels, such that the learning machine is expected to correctly classify unseen test examples. The best mapping of f that can be obtained is by minimizing the *expected error*, represented by the following expression (Vapnik, 1998):

$$R[f] = \int \frac{1}{2} |y - f(x)| \; dP(x, y) \tag{4.6}$$

where the component $|y - f(x)| \; dP(x, y)$ is known as the *loss function*. A further common loss function for example is $\big(y - f(x)\big)^2$, which is also known as the squared loss function.

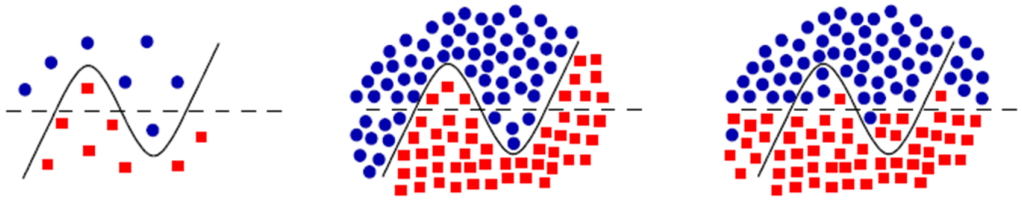


Figure 4.6: Overfitting phenomenon. The more complex function obtains a smaller training error than the linear function (left). But only with a larger data set it is possible to decide whether the more complex function really performs better (middle) or overfits (right)

Unfortunately, the expected error cannot be directly minimized, since the probability distribution function $P(x, y)$, from which the data is generated, is unknown. In order to estimate a function f that is close to the optimal one, an *induction principle* for risk minimization is therefore necessary. The most straightforward way is to approximate the minimum of the risk discussed in equation (4.7) by the minimum of the so called *empirical risk*, namely the measured *mean error rate* on the training set (Vapnik, 1998):

$$R_{emp}[f] = \frac{1}{l}\sum_{i=1}^{l}\frac{1}{2}|y - f(x_i)| \tag{4.7}$$

A small error on the training set does not necessarily indicate a high generalization ability. As anticipated in Section 4.1.3, this phenomenon is known as *overfitting*, as shown in Figure 4.6. In particular, as described in Müller *et al.* (2001), given a small training dataset as shown in the left of Figure 4.6, functions of f with higher degrees of complexity may result in smaller training errors. Nevertheless, only with a larger dataset, as the ones shown in the middle and right of Figure 4.6, it is understood that the decision on the right of Figure 4.6, reflects the true distribution more closely and does not overfit. One technique to avoid overfitting, is by restricting the complexity of the function f. This means that, a simple linear

function, describing most of the data is preferable to a complex function. These considerations, give rise to the problem of how to determine the optimal complexity of a function (Vapnik, 1995).

A specific technique introduced by Vapnik for controlling the complexity of a decision function is represented by the Structural Risk Minimization (SRM) principle (Vapnik, 1979). In order to understand how SRM works, it is necessary to introduce a non-negative integer h referred to as the *Vapnik-Chervonenkis dimension* or VC-dimension, which describes the complexity of a class of functions. In particular, the SRM measures the number of training points that can be separated for all possible labels using functions of that class. Once the concept of the VC-dimension is introduced, a nested family of function classes must be constructed (Vapnik, 1998):

$$\mathrm{F}_1 \subset \mathrm{F}_2 \subset \ \ldots \ \subset \mathrm{F}_k \tag{4.8}$$

whose VC-dimension satisfies:

$$h_1 \leq h_2 \leq \ \ldots \ \leq h_k \tag{4.9}$$

Then, suppose that the solutions of the empirical risk minimization problem:

$$f_1 \leq f_2 \leq \ \ldots \ \leq f_k \tag{4.10}$$

respectively belong to the function classes, $F_i, i = 1, \ldots k$. In that context, the SRM principle chooses the function f_i in the class F_i such that the right-hand side of the following bound on the generalization error is minimized (Vapnik, 1998):

$$R[f] \leq R_{emp}[f] + \sqrt{\left(\frac{h\left(\log\frac{2l}{h}+1\right)-\log\left(\frac{\eta}{4}\right)}{l}\right)} \tag{4.11}$$

where h is the VC-dimension of the function class under consideration, the square root term is called the *confidence term* and the bound holds with probability $1-\eta$ for any $0 \leq \eta \leq 1$.

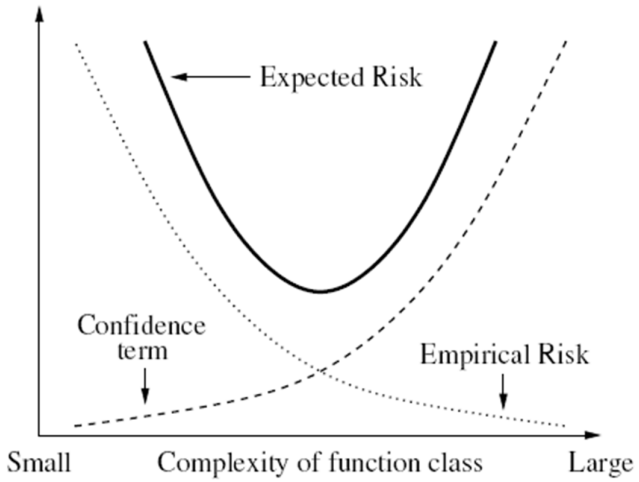


Figure 4.7: Schematic illustration of the bound in equation (4.11). The dotted line represents the empirical risk $R_{emp}[f]$. The dashed line represents the confidence term. The condintuous line represents the expected risk $R[f]$. The best solution is found by choosing the optimal tradeoff between the confidence term and the empirical risk $R_{emp}[f]$

Three aspects are of great interest in the above bound in equation (4.11). First, it is independent of $P(x,y)$, as it assumes only that the entire dataset (training and test set) is drawn independently according to some $P(x,y)$. Secondly, it is usually not possible to compute the left-hand side, namely the expected risk. Thirdly, the known VC-dimension h on the right-hand side is easily computable. This means that the selection of the learning machine which maps the input patterns to their class memberships by the function $f: \mathbb{R}^n \to \{\pm 1\}$ and minimizes the right-hand

side of equation (4.11), corresponds to the selection of the learning machine that achieves the lowest upper bound on the expected risk (Vapnik, 1995).

Minimization of the expected risk $R[f]$ is generally achieved by obtaining a small training error $R_{emp}[f]$ while keeping the function class to be small as possible. However, two extreme situations may arise. Firstly, a very small function class gives a disappearing square root term, but a relatively large training error. Secondly, a huge function class gives a disappearing empirical error, but a larger square root term. Nevertheless, from these considerations, it is evident that the best solution of the problem is usually in between, as shown in Figure 4.7. In other words, finding the minimum of the expected error actually means finding the right tradeoff between the accuracy obtained on that particular training set and complexity of the mapping created by the learning machine(Vapnik, 1995).

In practical problems the bound on the expected error discussed in equation (4.11) is neither easily comparable nor very helpful. A typical problem is that the VC-dimension of the class under consideration is either unknown or infinite. In such a case an infinite number of training data would be necessary. Nevertheless, the existence of bounds is important from a theoretical point of view, since it offers some deeper insights into the nature of learning machines.

4.2.2 Linking Statistical Theory to SVM

Linear learning machines such as the perceptron and SVM, as it will be clarified in the next section, use hyperplanes to separate classes in the feature space as shown in Figure 4.8, and can be represented mathematically as a function in the form (Vapnik, 1979):

$$y = \text{sign}(w \cdot x + b) \tag{4.12}$$

In Vapnik and Chervonenkis (1974) it has been demonstrated that the VC-dimension can be bounded in terms of another quantity referred to as a *margin*, namely the minimal distance of patterns from the hyperplane.

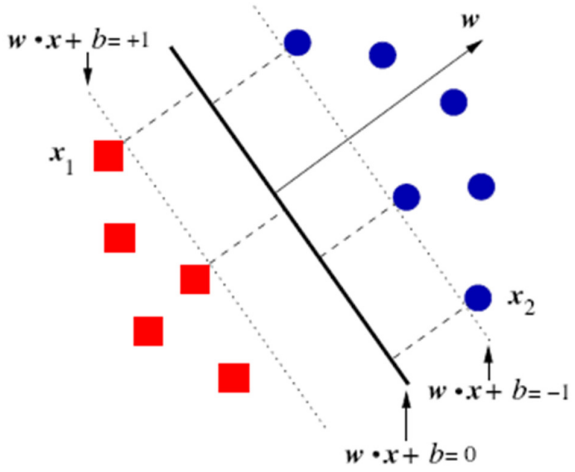


Figure 4.8: A hyperplane separating different patterns. The margin is the minimal distance between the pattern and the hyperplane, thus here the dashed lines

In Figure 4.8 the margin corresponds to the dashed lines. In particular, by rescaling w and b such that the points that are closest to the hyperplane satisfy the condition $|w \cdot x + b| = 1$, namely transforming the hyperplane to its canonical representation, it is possible to measure directly the margin as a function of w.

Consider, two patterns x_1 and x_2 belonging to two different classes and such that $|w \cdot x_1 + b| = +1$ and $|w \cdot x_2 + b| = -1$. Then the margin can be calculated as the distance between those two points along the perpendicular, namely as Vapnik and Chervonenkis (1974):

$$\frac{w}{\|w\|}(x_1 - x_2) = \frac{2}{\|w\|} \tag{4.13}$$

Then, it can be demonstrated that the inequality which links the VC-dimension of the class of separating hyperplanes to the margin uses the following expression:

$$h \leq \Lambda^2 R^2 + 1 \quad \text{and} \quad \|w\| < \Lambda \tag{4.14}$$

where R represents the radius of the smallest ball around the data. Due to the inverse proportionality between the margin and $\|w\|$ (in equation (4.13)), a small VC-dimension is obtained by requiring a large margin. On the other hand, a high VC-dimension is obtained by requiring a small margin. The bound described in equation (4.11) demonstrates that in order to achieve a small expected error it is necessary to keep both the training error and the VC-dimension small, then when working with linear learning machines, separating hyperplanes can be constructed such that they maximize the margin and separate the training patterns with as few errors as possible (Vapnik and Chervonenkis, 1974). As it will be demonstrated in the following section, this result forms the basis of the SVM learning algorithm.

4.2.3 Linear SVM

4.2.3.1 Separable Case

In order to introduce the SVM learning algorithm, the simplest case to deal with, is the *separable case* in which data is linearly separable. As it will be discussed in the following section, the most general case, namely, the *non-linear SVM* trained on non-separable data results in a similar solution. Suppose again that l training patterns are given (Vapnik and Chervonenkis, 1974):

$$(x_1, \dots, x_l) \qquad \text{with} \qquad x_i \in \mathbb{R}^n \quad \forall i = 1, \dots, l \tag{4.15}$$

together with the associated class labels representing the attended class membership:

$$(y_1, \dots, y_l) \qquad \text{with} \qquad y_i = \pm 1 \quad \forall i = 1, \dots, l \tag{4.16}$$

Assume that the patterns are linearly separable, namely they could be separated by a hyperplane $y = \text{sign}(w \cdot x + b)$ as the one shown in Figure 4.8. For such a learning machine, the conditions for classification without the training error rate are:

$$y_i(w_i \cdot x_i + b) \geq 1 \qquad i = 1,2, \dots, l \tag{4.17}$$

The aim of learning is thus finding w and b such that the expected risk is minimized. According to equation (4.11), one strategy that can be implemented, is to make the empirical risk zero by forcing w and b to result in a perfect separation between the two classes, while at the same time minimizing the complexity term of the VC-dimension h. For the case of a linear learning machine the VC-dimension h is bounded as described in equation (4.14), so it is possible to minimize the VC-dimension by minimizing the $\|w\|^2$ term, namely by maximizing the margin (Vapnik and Chervonenkis, 1974). The linear learning machine ensures the lowest expected risk is that, which gives an empirical risk zero, or perfect separation between the two classes represented by equation (4.17) and at the same time minimizes the VC-dimension or maximizes the margin between the two classes using the expression:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{4.18}$$

In order to solve this convex quadratic programming (QP) optimization problem, it is preferable to introduce a Lagrangian $\mathcal{L}$ (Vapnik and Chervonenkis, 1974):

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i (y_i (w_i \cdot x_i + b) - 1) \tag{4.19}$$

with the Lagrangian multipliers satisfying $\alpha_i > 0$, $i = 1,2, \ldots l$. The Lagrangian $\mathcal{L}$ has to be minimized with respect to w and b and to be maximized with respect to α_i. The conditions that at the saddle points the derivatives vanish are as follows:

$$\frac{\partial \mathcal{L}(w,b,\alpha)}{\partial b} = 0 \tag{4.20}$$

$$\frac{\partial \mathcal{L}(w,b,\alpha)}{\partial w} = 0 \tag{4.21}$$

leads to:

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{4.22}$$

$$w = \sum_{i=1}^{l} \alpha_i y_i x_i \tag{4.23}$$

By substituting equation (4.23) in equation (4.19), the dual quadratic optimization problem is obtained:

$$\max_{\alpha} \ \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \alpha_i \alpha_j y_i y_j \left(x_i \cdot x_j\right) \tag{4.24}$$

subject to the constraints:

$$\alpha_i > 0, \quad i = 1,2, \ldots, l \tag{4.25}$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{4.26}$$

Thus, solving the dual optimization problem, the Lagrangian multipliers $\alpha_i > 0$, $i = 1,2, \dots l$, need to express the specific w which solves equation (4.18). In particular, for each input pattern x, the following decision function will be applied (Vapnik, 1998):

$$f(x) = \text{sign}\left(\sum_{i=1}^{l} \alpha_i y_i \, (x \cdot x_i) + b\right) \tag{4.27}$$

The hyperplane given by the decision function in equation (4.27) is generally referred to as the *maximal margin hyperplane* of linear SVM.

4.2.3.2 Non-separable Case

When dealing with noisy data, it could happen that the patterns are not linearly separable. In such a situation it is impossible to keep the empirical error zero, therefore, it is necessary to find the best tradeoff between the empirical risk and the complexity term as discussed in equation (4.11). In order to relax hard-margin constraints, thus allowing classification errors, slack variables must be introduced (Cortes & Vapnik, 1995) such that:

$$y_i(w_i \cdot x_i + b) \geq 1 - \xi_i, \qquad \xi_i > 0 \qquad i = 1,2, \dots, l \tag{4.28}$$

In this case, the solution is found by minimizing the VC-dimension and an upper bound on the empirical risk, which represents the number of training errors. Thus the quantity to minimize is:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{l} \xi_i \tag{4.29}$$

In analogy to equation (4.18), finding a minimum of the first terms actually means minimizing the VC-dimension of the class of functions under consideration. On the other hand, the term $\sum_{i=1}^{l}\xi_i$ is an upper bound, representing the number of misclassifications on the training set, so finding a minimum for it results in minimizing the empirical risk. In this context, the SVM regularization parameter $C > 0$ determines the tradeoff between the complexity term and the empirical risk. As in the linearly separable case, the Lagrangian multipliers for the non-separable case are obtained by solving the following quadratic problem (QP) (Vapnik and Chervonenkis, 1974):

$$\max_{\alpha} \mathcal{L} \quad \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,k=1}^{l}\alpha_i\alpha_j y_i y_j \left(x_i \cdot x_j\right) \tag{4.30}$$

subject to the constraints:

$$0 \le \alpha_i \le C, \quad i = 1,2,\dots,l \tag{4.31}$$

$$\sum_{i=1}^{l}\alpha_i y_i = 0 \tag{4.32}$$

Thus the only difference of the non-separable case from the separable case is that the Lagrangian multipliers are upper bounded by the constant term, $C > 0$. The Karush-Kuhn-Tucker (KKT) conditions (Lasdon, 1970), state the necessary requirements for a set of variables to be optimal for an optimization problem, only those Lagrangian multipliers $\alpha_i, i = 1,2,\dots,l$ corresponding to a training pattern x_i which are either on the margin or inside the margin area are considered as non-zero. The KKT conditions assert that (Vapnik, 1995):

$$\alpha_i = 0, \quad \Rightarrow \quad y_i f(x_i) \ge 1 \quad \text{and} \quad \xi_i = 0 \tag{4.33}$$

$$0 < \alpha_i < C, \quad \Rightarrow \quad y_i f(x_i) = 1 \quad \text{and} \quad \xi_i = 0 \tag{4.34}$$

$$\alpha_i = C, \quad \Rightarrow \quad y_i f(x_i) \le 1 \quad \text{and} \quad \xi_i \ge 0 \tag{4.35}$$

These considerations reveal a fundamental property of SVM, namely that the solution found is sparse in α. This is crucial for computational time, since sparsity guarantees that the expansion discussed in equation (4.23) is calculated on the restricted number of patterns x_i corresponding to $\alpha_i > 0$, also known as *Support Vectors* (SVs) (Vapnik, 1995). The KKT conditions are also useful to compute the threshold b in equation (4.27). From equation (4.34), it follows that (Vapnik, 1995):

$$y_i\left(\sum_{j=1}^{l} \alpha_j y_j \left(x_i \cdot x_j\right) + b\right) = 1 \tag{4.36}$$

4.2.4 Non-linear SVM

SVM can afford to use more complex decision functions by remapping the input patterns into a higher dimensional space, in which the classification between the two classes can be performed by a separating hyperplane (Vapnik, 1998):

$$\Phi: \mathbb{R}^n \rightarrow \mathcal{H} \tag{4.37}$$

$$\mathrm{x} \rightarrow \Phi(x) \tag{4.38}$$

Suppose for example that some non-linearly separable patterns are given in two dimensions, as shown in Figure 4.9. By remapping them into a three-dimensional space of the second-order monomials (Vapnik, 1998):

$$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^3 \tag{4.39}$$

$$(x^1, x^2) \rightarrow \left((x^1)^2, \sqrt{2x^1}\, x^2, (x^2)^2\right) = (z^1, z^2, z^3) \tag{4.40}$$

the result is a linear hyperplane separating those patterns, as shown on the right side of Figure 4.9.

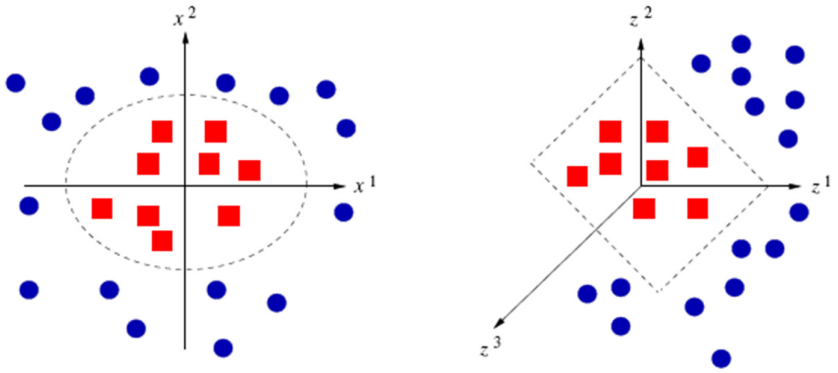


Figure 4.9: Non-linearly separable patterns in two-dimensions (left). By remapping them in a three dimensional space of the second order monomials (right) a linear hyperplane separating those patterns can be found

The SVM optimization problem generally involves the dot products of the training patterns x_i, as it is evident from equation (4.27). Therefore, the non-linear mapping $\Phi: \mathbb{R}^n \rightarrow \mathcal{H}$ that maps the patterns x_i into the new space known as a Hilbert space $\mathcal{H}$, does not require to be given explicitly. It is however necessary to specify the dot product of any of the two images $\Phi(x)$ and $\Phi(y)$ in $\mathcal{H}$ space through a $Kernel\ function\ (K)$ defined over $C \times C$, where C is a compact subset of $\mathbb{R}^n$ including the training and test patterns. The kernel function K can then be defined by the following expression (Vapnik, 1998):

$$K(x, y) \equiv \Phi(x) \cdot \Phi(y) \tag{4.41}$$

In order to assure that the kernel function definition in equation (4.41) is well posed, $K(x, y)$ must satisfy Mercer's conditions (Mercer, 1909). More specifically, $K(x, y)$ must be symmetric and continuous over $C \times C$. Once Mercer's conditions are satisfied, it is relatively easy to find a mapping Φ of the input patterns onto the Hilbert space $\mathcal{H}$ (Mercer, 1909). The maximal margin hyperplane can then be

represented in terms of the input patterns in $\mathbb{R}^n$, resulting in the following expression for the decision function (Vapnik, 1998):

$$f(x) = \text{sign}\left(\sum_{i=1}^{l} \alpha_i y_i K(x, x_i) + b\right) \tag{4.42}$$

Many kernel functions have been developed that can be used with SVMs. The kernels summarized in Table 4.1 (in equations (4.43) to (4.45)) satisfy Mercer's conditions (Mercer, 1909) and are commonly used with SVM. The sigmoidal kernel (in equation (4.43)) will only satisfy Mercer's condition for particular values of the *free parameters* (Burges, 1988), (Smola & Schölkopf, 2004), but this kernel has been used successfully in practice (Vapnik, 1995). The polynomial kernel of degree p (in equation (4.44)), is inhomogeneous such that. it allows the additive constant C to be larger than zero (Burges, 1988) for additional degrees of freedom (Boset *et al.*, 1992).

Table 4.1: Non-linear kernels commonly used to perform a dot product in a mapped feature space in the SVM formulation

Name	**Parameters**	**Kernel Function**	**Equation No.**
Polynomial	$c \in \mathbb{R}, p \in \mathbb{N}$	$K(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i} \cdot \mathbf{x_j} + c)^p$	(4.43)
Radial Basis Function (RBF) *Gaussian kernel*	$\gamma \in \mathbb{R}$	$K(\mathbf{x_i}, \mathbf{x_j}) = e^{-\gamma \|\mathbf{x_i} - \mathbf{x_j}\|^2}$	(4.44)
Sigmoidal	$\kappa \in \mathbb{R}, \delta \in \mathbb{R}$	$K(\mathbf{x_i}, \mathbf{x_j}) = \tanh(\kappa(\mathbf{x_i} \cdot \mathbf{x_j}) - \delta)$	(4.45)

The Radial Basis Function (RBF) kernel (in equation (4.45)) also known as the *Gaussian* kernel is the most widely used kernel with SVMs. The RBF kernel is translation invariant, this means that, $K_\gamma(\mathbf{x_i}, \mathbf{x_j}) = K_\gamma(\mathbf{x_i} - \mathbf{x_j})$ has an infinite

number of dimensions (Vapnik, 1995), (Burges, 1988). A significant advantage of the RBF kernel is that it adds only a single free parameter $\gamma > 0$, which controls the width of the RBF kernel as $\gamma = 1/2\sigma^2$, where σ^2 is the variance of the resulting Gaussian hypersphere. The RBF kernel has been shown to perform well in a wide variety of practical applications, such as in Degroeve *et al.* (2005), Hsu *et al.* (2003) and Wang *et al.* (2005).

Using the RBF kernel with SVM, there are two SVM hyperplane parameters that need to be determined in the SVM model which are: C and γ. In order to get good generalization ability, a validation process must be performed in order to decide the optimum values of these parameters. The procedure for SVM hyperparameter optimization presented by Hsu *et al.* (2003), which is known as the *Grid Search* method is as follows:

1. Consider a grid space of (C, γ) with $\log_2 C \in \{-1, -2, \dots, 20\}$ and $\log_2 \gamma \in \{-20, -19, \dots, 1\}$.
2. For each SVM hyperparameter pair (C, γ) in the search space, perform 10-fold CV on the training set.
3. Choose the parameters (C, γ) that lead to the highest (optimum) CV accuracy and lowest error, and use them to build the SVM classification engine (trained model).

4.2.5 Implementation of SVM

As discussed in Sections 4.2.3 and 4.2.4 previously, the training of SVM requires the solution of a convex quadratic programming (QP) optimization problem. This task is very difficult to implement by average engineers and the training

algorithms that use numerical QP are slow, especially for large size problems. In the past decade, a number of researchers introduced new learning algorithms that use faster and simpler-to-implement methods for solving QP problems in SVMs. These methods include the: Newton method, Quasi Newton method, Kernel Adatron (KA) (Frieß *et al.*, 1998) and Sequential Minimal Optimization (SMO) (Platt, 1998 & 1999a).

This research focuses on SMO because other optimization methods have proven to be slower (Platt, 1999b) on large of datasets and also since the LIBSVM (Chang & Lin, 2010) library used for implementing SVM in this research also uses SMO. The following section gives a brief description of the SMO algorithm and its implementation for SVM.

4.2.5.1 Sequential Minimal Optimization

As the training of SVMs under normal circumstances requires a lot of time for calculation of the Kernel matrix, the training time increases tremendously when a large number of training samples are present, resulting in a bigger Kernel matrix. In order to solve this problem, SMO deals with the large QP problems, by breaking (decomposing) the problem into a series of smaller QP problems.

The Sequential Minimal Optimization (SMO) algorithm was first introduced by John C. Platt in 1998 (Platt, 1998 & 1999a). The main concept of SMO is to break large QP problems into smaller QP problems. More specifically, in SMO a minimal subset of only two training samples can be optimized on each training iteration. This is because the smallest number of Lagrangian multipliers that can be used for optimization at each step, is two. Thus, in this way each small QP problem is solved

analytically without the need of performing time consuming numerical optimization, which makes implementation of SMO easy and simple.

Table 4.2: Summarized procedure of the SMO algorithm

Step No.	Procedure
1	Choose the first Lagrange multiplier to be a KKT violator.
2	Choose the second Lagrange multiplier using heuristics.
3	Update the second Lagrange multiplier via: $\alpha_2^{new} = \alpha_2 + \frac{y_2(E_1 - E_2)}{\kappa}$
4	Clip the multiplier α_2^{new} to $\alpha_2^{new,clipped}$
5	If the multiplier does not change, go back to Step 1.
6	Update the first Lagrange multiplier.
7	Update the error-cache.
8	If all Lagrange multiplier fulfill KKT conditions, stop; else go to Step 1.

At every step of SMO, two Lagrange multipliers are selected for optimization and after their optimal values are found given that all the other multipliers are fixed, the SVM is updated accordingly. In SMO, the two training samples are selected using a heuristic method, and then the two Lagrange multipliers are solved analytically. The SMO algorithm is summarized and presented in Table 4.2 (Platt, 1998 & 1999a).

The main advantage of SMO is that it uses only two training samples at every step and avoids the computation of a Kernel matrix. Due to this reason, SMO requires a smaller amount of memory and can handle very large training sets compared to other optimization techniques (Platt, 1998 & 1999a) such as the *chunking* algorithm (Vapnik, 1979).

4.3 Artificial Neural Networks

In order to estimate the performance of the SVM and perform a comparative research as outlined in the fifth research objective in Section 1.2, different machine learning algorithms other than SVM are evaluated in this research. Since ANNs have similar structure to that of SVMs, so they are used for the purpose of model comparison.

In this research two ANN based approaches are identified, i.e. a *traditional* and a *modern* ANN approach, namely the Back-Propagation Neural Network (BPNN) (Rumelhart *et al*., 1986) and the Extreme Learning Machine (ELM) (Huang *et al*., 2006a) respectively. The Online-Sequential Extreme Learning Machine (OS-ELM) (Liang *et al.*, 2006) is a recently proposed variant of the ELM (Huang *et al*., 2006a), which overcomes the limitations of standard ELM. The BPNN (Rumelhart *et al*., 1986) and the OS-ELM (Liang *et al.*, 2006) as discussed in the following sections are adopted in this research.

4.3.1 Back-Propagation Neural Network (BPNN)

Back-propagation (BP) also referred to as "propagation of error" is a common method used for teaching ANNs on how to perform given tasks. The BP algorithm was first presented by Paul Werbos in 1974, however, it wasn't until 1986 that it gained importance in the field of machine learning research (Rumelhart *et al*., 1986). The BP algorithm is most suitable for use with feed-forward networks. Feed-forward networks are those type of networks that have no feedback.

The BP is a supervised learning method, which is an implementation of the Delta rule. The delta rule in BP typically requires a teacher (or trainer) that knows, or

can calculate, the desired output for any possible input. In the BP algorithm, the errors propagate backwards from the output nodes of the network to the inner nodes of the network. So the BP algorithm is generally used to calculate the gradient error of the network. The gradient of the error is typically used in a gradient descent algorithm to determine the weights that minimize the training error (Rumelhart *et al*., 1986). BP networks are typically multi-layer ANNs, usually with an input layer, one or more hidden layers and an output layer. For the hidden layer neurons to serve any useful purpose, they must have non-linear activation (or transfer) functions. The most common non-linear activation functions include the: log-sigmoid, tan-sigmoid, Gaussian and softmax transfer functions.

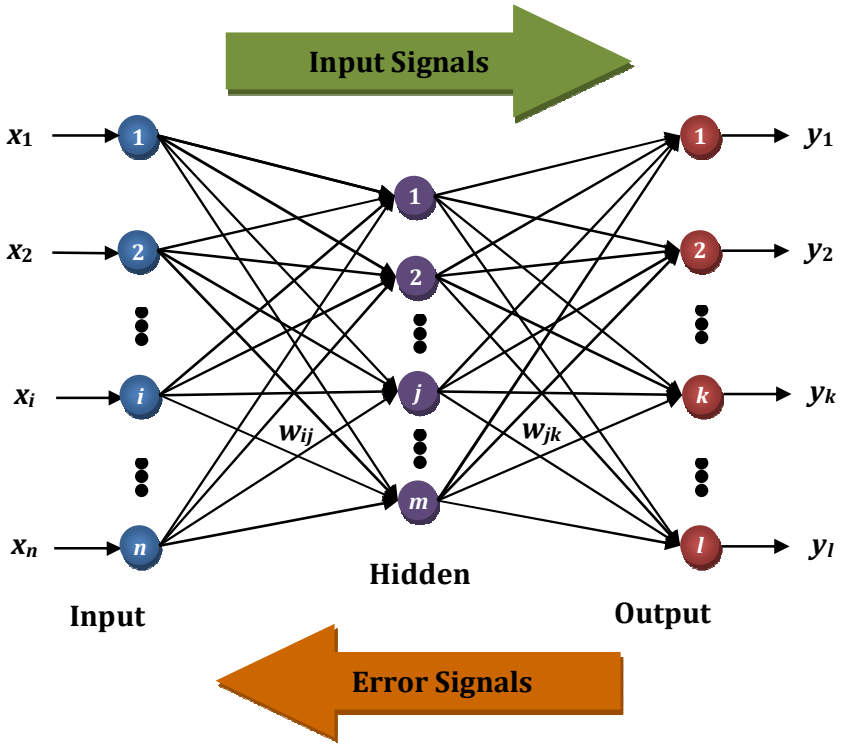


Figure 4.10: General architecture of a Back-Propagation Neural Network (BPNN)

In a Back-propagation Neural Network (BPNN), learning is formulated as follows. Firstly, a training pattern is presented to the input layer of the BPNN. The network

propagates the input pattern from layer to layer until the output pattern is generated by the neurons in the output layer. If the output pattern is different from the desired output, an error is calculated. This error is then propagated backwards through the network to the input layer. As the error is propagated backwards, the weights connecting the neurons are adjusted by the BP algorithm. Figure 4.10 shows a typical architecture of a BPNN (Wen *et al.*, 2000). The following steps illustrate the implementation of the BP training algorithm.

Step 1 — Network Initialization

All the weights and threshold levels of the network are set to small random numbers uniformly.

Step 2 — Activation Function

The BPNN is activated by applying inputs $x_1(k), x_2(k), \dots, x_n(k)$ and desired outputs $y_1(k), y_2(k), \dots, y_n(k)$ as shown in Figure 4.10. Then the actual output of the neurons in the hidden layer is calculated using the following transfer function:

$$y_j(k) = logsig\left[\sum_{i=1}^{n} x_i(k) \cdot w_{ij}(k) - \theta_j\right] \tag{4.46}$$

where n is the number of inputs of neuron j in the hidden layer and $logsig$ represents the log-sigmoid activation function.

Step 3 — Weight Adjustment

The weights in the BPNN are updated and the errors associated with output neurons are propagated backward.

Step 4 — Iteration

The value of k is increased by one, and Step 2 is repeated again. The iterations continue until the error becomes zero or a desired performance goal is met, whichever comes first.

As the BP algorithm is a supervised machine learning algorithm, it is applicable for the task of classification and regression. Many complex activation functions have been proposed by researchers in the past few decades to overcome the limitation of standard activation functions, and the BP algorithm is applicable to all. For further reading on the BP training algorithm see Rumelhart *et al*. (1986).

4.3.2 Online-Sequential Extreme Learning Machine (ELM)

The Extreme Learning Machine (ELM) was proposed by Huang in 2006 (Huang *et al*., 2006a) for Single Layer Feed-forward Networks (SLFNs), which states to produce superior performance (Huang and Chee-Kheong, 2004), (Huang *et al*., 2006b), (Huang *et al*., 2006c) compared to other machine learning algorithms. The ELM is one algorithm amongst the supervised batch learning algorithms that uses a finite number of input and output samples for training. The ELM algorithm is claimed to be extremely fast in its learning speed and has better generalization performance compared to conventional learning algorithms (Huang *et al*., 2004).

The ELM is a modern learning algorithm for SLFNs that works efficiently for classifications, function approximations and online prediction problems. Moreover, the ELM can work well for a variety of types of applications. Typically a SLFN has three input parameters: (i) input weight w_i, (ii) hidden neuron biases b_i, and (iii) output weight β_i. While traditional learning algorithms of SLFNs have to tune these

parameters, the ELM randomly generates the input weight w_i, the hidden neuron biases b_i and then calculates the output weight β_i. Thus, for SLFNs trained using the ELM algorithm, no further learning is required.

Given N arbitrary distinct samples (x_i, t_i), where $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in \mathbb{R}^m$, a SLFN with $\widetilde{N}$ hidden neurons and transfer function $g(x)$ can be represented by:

$$\sum_{i=1}^{\widetilde{N}} \beta_i g(w_i \cdot x_j + b_i) = o_j, \quad j = 1,2, \dots, N \tag{4.47}$$

where w_i represents the weight vector connecting the input neurons and the i^{th} hidden neuron i, b_i is the threshold of the i^{th} hidden neuron and β_i is the weight vector connecting the i^{th} hidden neuron and the output neurons. In equation (4.47) $w_i \cdot x_j$ represents the inner product of w_i and x_j. If the network is able to approximate these N samples with zero error, then $\sum_{j=1}^{N} \|o_j - t_j\| = 0$,, there exists β_i, w_i, b_i such that, $\sum_{i=1}^{\widetilde{N}} \beta_i g(w_i \cdot x_j + b_i) = t_j$ where $j = 1,2, \dots, N$. Thus, the above equations can be represented as $H\beta = T$, where:

$$H(w_1, \dots, w_{\widetilde{N}}, b_1, \dots, b_{\widetilde{N}}, x_1, \dots, x_{\widetilde{N}}) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_{\widetilde{N}} \cdot x_1 + b_{\widetilde{N}}) \\ \vdots & \cdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_{\widetilde{N}} \cdot x_N + b_{\widetilde{N}}) \end{bmatrix}_{N \times \widetilde{N}} \tag{4.48}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\widetilde{N}}^T \end{bmatrix}_{N \times m} \quad \text{and} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_{\widetilde{N}}^T \end{bmatrix}_{N \times m} \tag{4.49}$$

As discussed in Huang and Babri (1998), H is the hidden layer output matrix of the ANN, with the i^{th} column of H being the i^{th} hidden neuron output for the inputs $x_1, x_2, \dots, x_n$. Based on the previous work of Huang, matrix H is considered to be square and invertible only if the number of hidden neurons equals the number of distinct training samples $\widetilde{N} = N$. Satisfying the above condition for matrix H indicates that SLFN can approximate the training samples with almost a zero error. Typically the number of hidden neurons is much lower than the number of distinct training samples, $\widetilde{N} \ll N$. In equation (4.48) H is a non-square matrix, which indicates that there may not exist $w_i, b_i \beta_i (i = 1, \dots, N)$ such that $H\beta = T$. Thus, a specific set of $\widehat{w}_i, \hat{b}_i \hat{\beta}_i (i = 1, \dots, \widetilde{N})$ needs to be found so that:

$$\left\| H\left(\widehat{w}_1, \dots, \widehat{w}_{\widetilde{N}}, \hat{b}_1, \dots, \hat{b}_{\widetilde{N}}\right)\hat{\beta} - T \right\| = \min_{w_i, b_i, \beta_i} \| H(w_1, \dots, w_{\widetilde{N}}, b_1, \dots, b_{\widetilde{N}})\beta - T \| \quad (4.50)$$

which is equivalent to minimizing the cost function,

$$E = \sum_{j=1}^{N} \left(\sum_{i=1}^{\widetilde{N}} \beta_i g(w_i \cdot x_j + b_i) - t_j \right)^2 \quad (4.51)$$

Huang in (Huang *et al.*, 2006b) and (Huang, 2003) discussed that the hidden neuron parameters in the ELM need not be tuned, as the matrix H converts the data from non-linear separable cases to high-dimensional linear separable cases. Furthermore, Huang in (Huang *et al.*, 2004) showed that the input weights and hidden neurons need not to be tuned and can be randomly selected and then fixed. Thus, for fixed input weights and the hidden layer biases (kernel parameters), training a SLFN is equivalent to finding a least squares solution $\hat{\beta}$ for the linear system, $H\beta = T$.

In order to handle online applications a variant of the ELM typically known as the Online-Sequential Extreme Learning Machine (OS-ELM) was introduced by Liang *et al*. (2006). The OS-ELM was proposed to overcome the limitations of ELM as developed by Huang (Huang *et al.*, 2004). Since the ELM algorithm belongs to supervised batch learning algorithms, this prohibits its further application. As in the real world, training data may arrive either chunk-by-chunk or one-by-one, therefore, an online-sequential learning is most suitable to cater for such variations. The OS-ELM was originally developed for SLFNs with additive or RBF hidden nodes in a unified framework, thus it can handle both additive neurons and Radial Basis Function (RBF) nodes. Unlike other sequential learning algorithms that require many parameters to be tuned, the OS-ELM only requires the number of hidden layer neurons to be specified. Huang *et al*. (2004) proposed that in order to determine an optimum value for the parameter n, an iterative approach needs to be applied, which is as follows:

Start with an initial value of $n = 20$

Increment n by 20 on each iteration.

Calculate the training accuracy of the model using for n.

Stop iterations when $n = 200$

The OS-ELM algorithm as proposed by Liang *et al*. (2006) consists of two major phases, namely the, *initialization* (or boosting) phase and the *sequential-learning* phase. In the initialization phase, the number of data required should be equal to the number of hidden layer neurons. The initialization phase trains the SLFN using the OS-ELM method given by some batch of training data. This data is discarded once the process is complete. Following the initialization phase, in the learning

phase the OS-ELM learns the training data using a chunk-by-chunk procedure. All the training data is discarded once the learning procedure involving the data is complete.

The OS-ELM batch training algorithm, provides a faster learning capability as compared with traditional machine learning techniques. Unlike other popular learning machines, only a small amount of human involvement is required in implementing the OS-ELM. Except for the number of the hidden neurons (insensitive to OS-ELM), no other network parameters need to be optimized by the users, since the OS-ELM algorithm chooses the input weights randomly and analytically determines the output weights itself Huang *et al*. (2006b).

4.4 Recursive Feature Elimination

A challenging task in pattern classification problems, namely machine learning, is to reduce the dimensionality n of the feature space by finding a restricted number of features yielding good classification performance. In recent years, a lot of work has been done in the direction of feature selection (Kohavi and John, 1997), (Kearns *et al.*, 1997). Feature elimination has known to be a fundamental process in order to reduce the computation time required to solve pattern classification problems and to improve the classification performance of the learning machine. The *curse of dimensionality* from statistical theory point of view, asserts that the difficulty of a prediction problem increases with the dimension n of the feature space, as in principle, exponentially many patterns are required to sample the space properly.

The SVM is introduced in this chapter as a tool for the solution of pattern classification problems. Next, new aspects of the applicability of SVMs in knowledge discovery and data mining, namely *feature selection* will be discussed. The reason for this is that SVMs are known to be very effective for discovering informative attributes of the dataset, namely critically important features. To serve this purpose, feature selection methods, namely, Recursive Feature Elimination (RFE) will be discussed.

Since, feature ranking is the first task addressed towards the elimination of unimportant features, this forms the basis for using Non-linear SVM (in Section 4.2.4) for the purpose of pattern classification. The algorithms presented in the following sections are usually known for combining SVM-based RFE using various strategies such as the F-score and the Random Forest (RF) (Chen and Lin, 2006).

4.4.1 Feature Ranking Using F-score

Feature ranking methods define the importance of each single feature according to its contribution to the learning machines' predictive accuracy. The final aim is to obtain a ranked list of features from which the features having an important contribution to the model can be selected, whereas features having a smaller contribution can be eliminated. Thus, feature ranking eliminates all those features which are useless for discrimination purposes, or at least represent noise.

Several methods evaluating how well individual features contribute towards a binary classification (two-class) problem have been indicated in the literature. For instance, Golub *et al*. (1999) used the following correlation coefficient $\left(r^i\right)$ as the feature ranking criterion:

$$r^i = \frac{\mu^i(p^a) - \mu^i(n^a)}{\sigma^i(p^a) + \sigma^i(n^a)} \tag{4.52}$$

where μ^i and σ^i are respectively the mean and the standard deviation of the feature i for all the patterns whose class is positive (p^a) or negative (n^a). Large positive r^i values represent a strong correlation with the class p^a, whereas large negative r^i values represent strong correlation with class n^a. Then, by selecting an equal number of features with positive and negative correlation coefficients, the two classes can be represented. Another approach, as described by Furey *et al.* (2000) used the absolute value $|r^i|$, whereas the authors in Pavlidis *et al.* (2001) used the following correlation coefficient:

$$r^i = \frac{\left(\mu^i(p^a) - \mu^i(n^a)\right)^2}{\left(\sigma^i(p^a) + \sigma^i(n^a)\right)^2} \tag{4.53}$$

An important drawback characterizing these feature ranking techniques (in equations (4.52) and (4.53)) is that, they rely on the implicit orthogonality assumptions that they make. In fact, each correlation coefficient (r^i) is computed by using only the information on that single feature, thus without taking into account the mutual information between features. This is a major problem, since features are typically correlated with each other, as the case with pattern classification problems. In order to overcome this problem, it is necessary to work with *multivariate learning machines*, namely the learning machines which are optimized during the training phase to handle multiple features simultaneously. SVM for example, is a typical multivariate learning machine (Chen & Lin, 2006).

The F-score is a simple technique which estimates the discrimination of two sets of real numbers. For a given number of training vectors x^k, $k = 1,2,\dots,m$ if the number of positive and negative instances are n_+ and n_-, respectively, then the F-score of the i^{th} feature is defined by the following expression (Chen & Lin, 2006):

$$F(i) = \frac{\left(\bar{x}_i^{(+)} - \bar{x}_i\right)^2 + \left(\bar{x}_i^{(-)} - \bar{x}_i\right)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}\left(\bar{x}_{k,i}^{(+)} - \bar{x}_i^{(+)}\right)^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}\left(\bar{x}_{k,i}^{(-)} - \bar{x}_i^{(-)}\right)^2} \tag{4.54}$$

where $\bar{x}_i$, $\bar{x}_i^{(+)}$, $\bar{x}_i^{(-)}$ are the average of the i^{th} feature of the positive and negative datasets, respectively. Similarly in equation (4.54), $\bar{x}_{k,i}^{(+)}$ is the i^{th} feature of the k^{th} positive instance and $\bar{x}_{k,i}^{(-)}$ is the i^{th} feature of the k^{th} negative instance. The numerator in equation (4.54) represents the discrimination between the positive and negative sets whereas the denominator represents the one within each of the two sets. In practice it is considered that, the larger the F-score is, the more likely this feature is more discriminative. Since the F-score is a simple and effective technique, the procedure for selecting optimum features (high F-scores) using the SVM is summarized in the following steps (Chen and Lin, 2006):

1. Calculate the F-score of every feature.
2. Pick possible thresholds (using the naked eye) to cut low and high F-scores.
3. Then for each threshold, do the following:
    a) Drop features with F-scores below this threshold.
    b) Split the training data into two random sets: X_{train} and X_{valid}.
    c) Set X_{train} be the new training data. Apply the technique discussed in Section 4.2.4, to obtained a binary SVM classifier; use the classifier to predict X_{valid}.

d) Repeat the steps above 10 times and then calculate the average validation error for each trial; perform 10-fold CV.

4. Choose the best threshold, i.e., the threshold with the lowest average validation error; lowest 10-fold CV accuracy.
5. Drop features with F-score below the selected threshold. Then apply the Non-linear SVM in Section 4.2.4.

In the above procedure, possible thresholds can be determined by the human eye. This step can be automated by gradually adding high F-score features, until the validation accuracy of the model decreases.

4.4.2 SVM-RFE Using Random Forest

The F-score feature ranking technique discussed for SVM-RFE in Section 4.4.1 (in equation (4.54)) is concerned with the removal of one feature at a time. Starting with a large number of features, when the goal is to obtain a small subset of relevant features, it is necessary to remove more than one feature at a time, which can be accomplished using filtering. Filtering can be accomplished by applying the Random Forest (RF) technique (Svetnik *et al.*, 2004).

The Random Forest (RF) is a typical classification tool, but also provides feature importance (Breiman, 2001). In the RF technique, a forest contains many decision trees which constructed using instances with randomly sampled features. So, the prediction using the RF is made by a majority vote of decision trees. In order to obtain feature importance, the training set is split into two sets. Training using the first set and predicting using the second set, an accuracy (a_i) can be obtained. Also, the values of the j^{th} feature are randomly permuted in the second set to obtained

another accuracy (a_j). Thus, the difference between these accuracies (a_i and a_j) shows the importance of the j^{th} feature (Chen & Lin, 2006).

In practice the RF technique is known to have a high computational time. Thus, before using the RF, a subset of features are selected at first using F-score feature ranking technique indicated in Section 4.4.1. The SVM-RFE technique applied in this research for the purpose of feature selection is referred to as "*F-score + RF + SVM*", which is summarized in the following steps (Chen & Lin, 2006):

1. F-score
    a) Consider the subset of features obtained using the F-score feature ranking technique in Section 4.4.1.
2. Random Forest (RF)
    a) Initialize the RF dataset so as to include all training instances with the subset of features selected from Step 1(a). Then use the RF technique to obtain the rank of features.
    b) Next, use the RF as a classification engine in order to perform 10-fold CV on the working set.
    c) Then update the working set by removing half of the less important features and go to Step 2(b). Stop if the number of features is small.
    d) Amongst the various feature subsets chosen above, select the feature subset with the lowest 10-fold CV error.
3. Support Vector Machine (SVM)
    a) Apply the technique discussed in Section 4.2.4, to obtained a binary SVM classifier;

4.5 Summary

An overview of pattern recognition was presented in this chapter, with particular emphasis on a specific machine learning technique, namely, the Support Vector Machine (SVM). SVMs are used intensively in this research. The reason for using SVM as the main machine learning technique for this research is discussed in Sections 1.3 and 3.3.5. Section 4.1 presented some introductory notions regarding the theoretical concepts of learning machines. Section 4.2 introduced the fundamental concepts of the statistical learning theory and presented the mathematical formulation of the SVM developed by Vapnik (1998) which describe the statistical aspects of automated machine learning. Towards the end of this chapter, Section 4.3 presented the theoretical concepts of ANNs whereas Section 4.4 discussed a Recursive Feature Elimination (RFE) technique used for the selection of the optimal subset of texture features for the learning machine (SVM).

CHAPTER 5

FRAMEWORK MODELING

5.0 Overview

This chapter presents modeling of the framework (system) proposed in Chapter 1 for the classification of benign and malignant abnormalities in digital mammograms. As discussed in Chapter 3, the proposed framework is composed of two main techniques, namely, image processing and machine learning. The image process and machine learning techniques identified in Section 3.3, are applied in the proposed system in this chapter, which are discussed in Section 5.1 and 5.2. The modeling of the system consists of three main stages, namely: Mammogram Image Processing (Section 5.3), Texture Feature Extraction and Selection (Section 5.4) and Classification Engine (Section 5.5). Sections 5.3 through 5.5 describe each stage in detail during the development of the proposed framework.

5.1 Proposed Framework

As discussed in Section 1.4, a prototype framework for the computerized detection of breast cancer from digital mammograms is presented in Figure 1.4. The modeling of this framework consists of three main stages, namely: Mammogram Image Processing (see Section 5.3), Texture Feature Extraction and Selection (see Section 5.4) and Classification Engine (see Section 5.5) as indicated in Figure 5.1.

In the first stage, image processing techniques and algorithms are applied on the digital mammographic images for the purpose of image preprocessing (see Section

3.3.1) and image segmentation (see Section 3.3.2). In this research, mammogram preprocessing includes: noise removal (using *Median Filtering*), background suppression (using *Global Thresholding*) and artifact/wedge and label suppression (using *Morphological Operations*) as shown in Figure 5.1.

During the mammogram segmentation stage, first the breast profile is optimally segmented from the background (using *Contrast Enhancement*) so that the breast-skin edges are retained in the breast profile (tissue). Secondly, the pectoral muscle is suppressed (using *Seeded Region Growing*) from the mammograms, since it may bias the procedures in the detection of malignant and benign tumors (Nicolaou *et al.*, 2008), (Xu *et al.*, 2007), (Mirzaalian *et al.*, 2007). The ROIs (abnormal regions) in the segmented mammogram images are extracted using the *Ground Truth* (GT) data from radiologists' diagnosis of the mammographic datasets, as discussed in Section 5.2.1.

In the second stage, texture features analysis is performed using Gray Level Co-occurrence Matrices (GLCMs), which are employed in this research to compute the texture features (see Section 3.3) from the ROIs (abnormal regions). Standard GLCM texture descriptors proposed by Haralick *et al*. (1973) and Haralick (1979) are used with other texture descriptors proposed by Soh and Tsatsoulis (1999) and Clausi (2002). The optimal subset of texture features is selected using the SVM-Recursive Feature Elimination (SVM-RFE) technique (using *F-score* + *Random Forest* + *SVM*) technique, as discussed in Section 4.4.

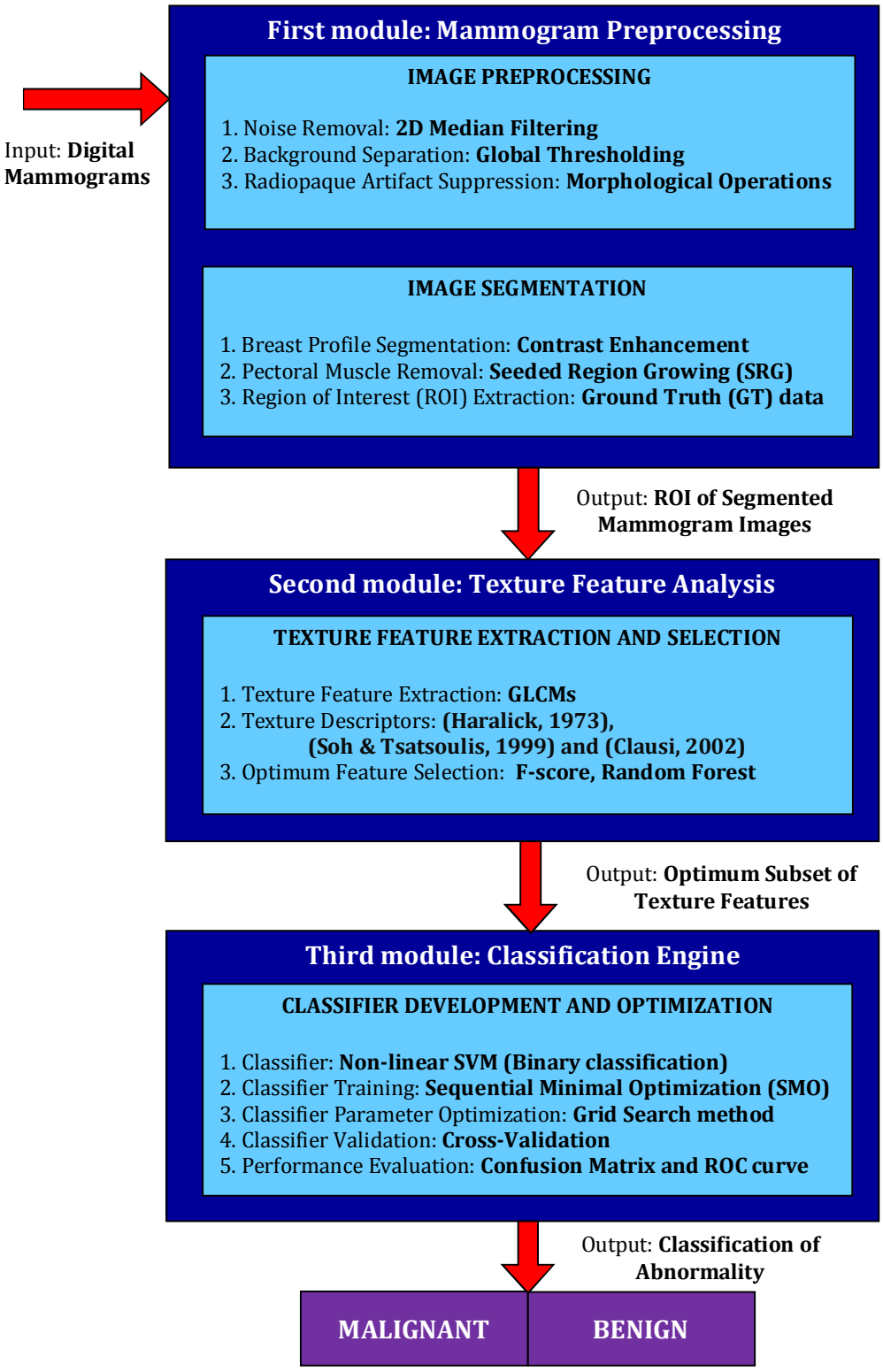


Figure 5.1: Flowchart of the proposed computerized breast cancer detection framework

The last stage in Figure 5.1 uses the optimal subset of texture features to construct a classification engine (classifier) using the non-linear SVM approach presented in Section 4.2.4. The SVM classifier used in this research performs binary classification between malignant and benign ROIs. The SVM is implemented using the Sequential Minimal Optimization (SMO) algorithm (see Section 4.2.5). The memorization capability of the SVM classifier is estimated using the Cross-Validation (CV) approach discussed in Section 4.1.3.2. The *Grid-Search* method presented in Section 4.2.4 is used to optimize (fine-tune) the hyperplane parameters for the non-linear SVM. The accuracy of the proposed classification engine is evaluated using a confusion matrix (in Table 3.1) such as the *sensitivity*, *specificity*, *FPF* and *AUC* computed for a medical diagnosis test, as discussed in Section 3.3.6. The output of the proposed system classifies the tested samples (ROIs) as malignant or benign.

5.2 Research Methodology and Implementation

The framework proposed in this thesis for the modeling of a computerized breast cancer detection system (in Figure 5.1) is indicated in Figure 5.2. The first three stages in Figure 5.2 will be discussed in this chapter, and the fourth stage which presents the testing results of the modeled framework, are discussed in Chapter 6. In this research, the image processing and texture analysis techniques are applied using the MATLAB Image Processing and Statistics Toolboxes. The SVM is applied in the proposed framework using the LIBSVM library (Chang & Lin, 2010).

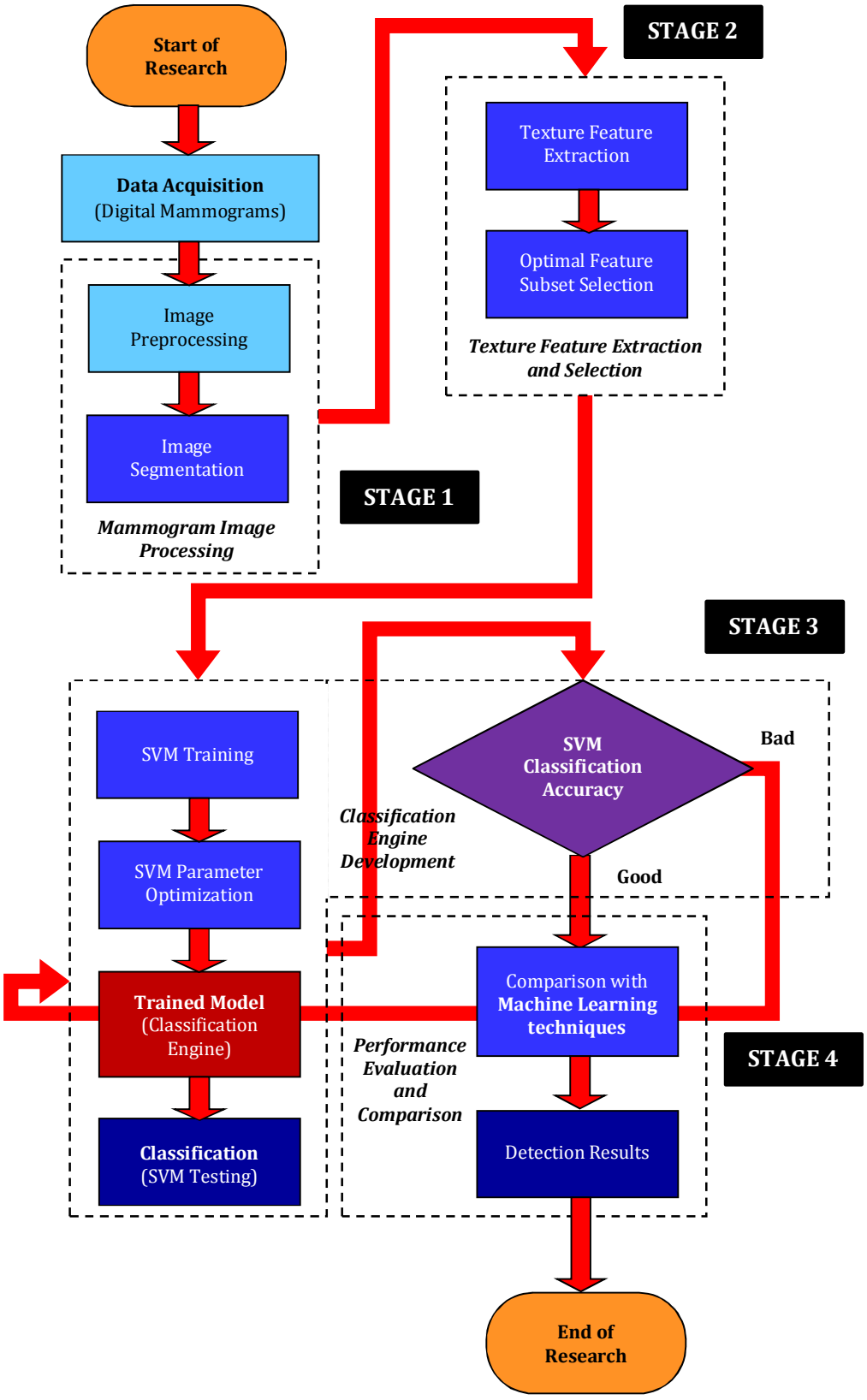


Figure 5.2: Flowchart of the research framework

5.2.1. Data Acquisition

Digital mammograms are used as the standard inputs into the proposed framework, as indicated in Figure 1.4 and Figure 5.1. The data used in this research is obtained from two distinct sources:

1. Mammographic images obtained from the University of Malaya Medical Centre (UMMC), Kuala Lumpur.
2. Mammography dataset obtained from the Mammographic Image Analysis Society (MIAS) database (Suckling *et al*., 1994).

The mammography images obtained from the UMMC are used in this research as the *local dataset*. In order to generalize the accuracy of the proposed system in better terms, a second dataset, namely the MIAS database (Suckling *et al*., 1994), is used in this research as the *external dataset*. Both datasets used in this research consist of the mediolateral oblique (MLO) view of the mammograms.

Digital mammograms in the local dataset were acquired for Malaysian patients treated at the Department of Radiology at the University Malaya Medical Centre (UMMC) (Dept. of Biomedical Imaging at UMMC, 2010) in collaboration with the Faculty of Medicine at the University of Malaya (Faculty of Medicine at UM, 2010).

The MIAS database of mammograms (Suckling *et al*., 1994) is a well-known published image database of 322 digital mammograms from the Mammographic Image Analysis Society, United Kingdom. The mammograms in the MIAS database are taken from the United Kingdom National Breast Screening Programme and this dataset has been cited in many peer reviewed research articles (Ferrari *et al*.,

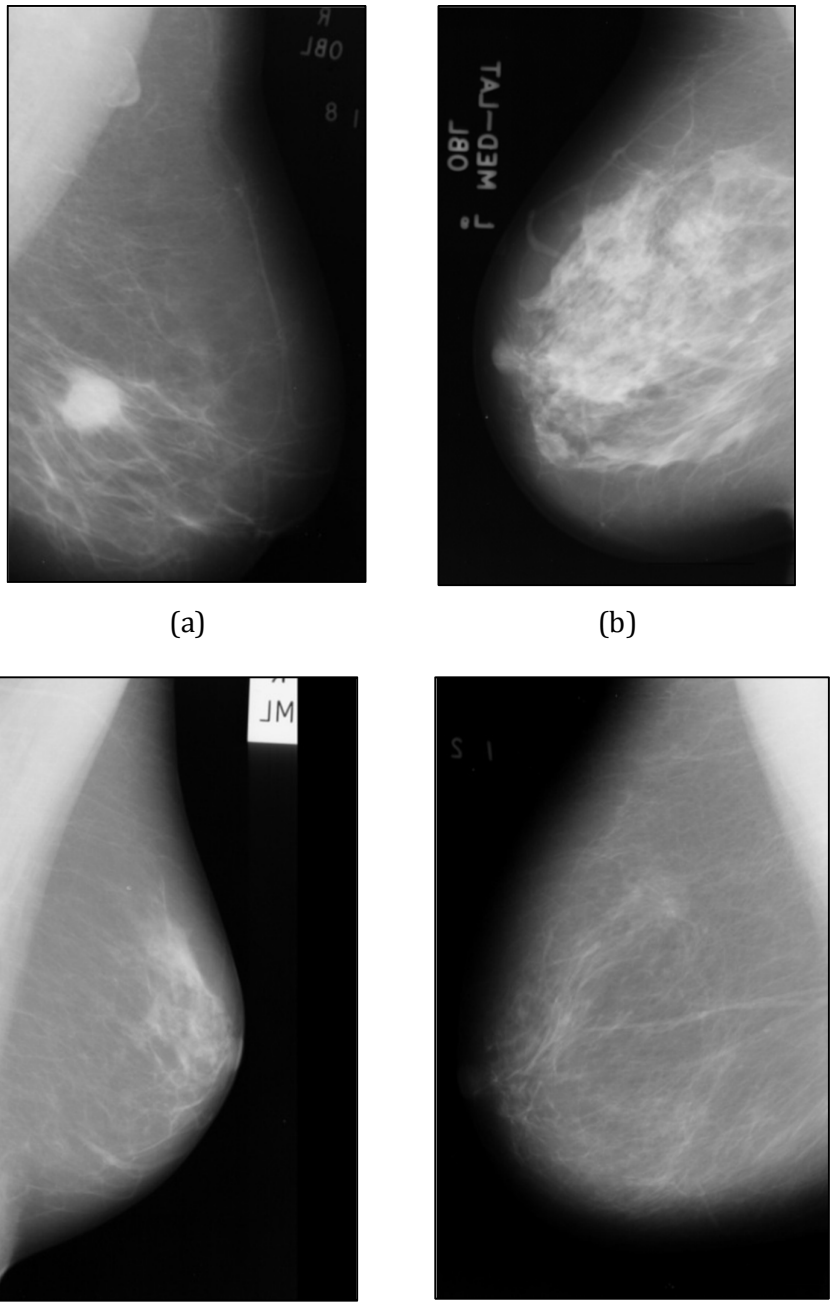
(2004, 2001)), (Selvan *et al*., 2006), (Dua *et al*., 2009), (Özekes *et al.*, 2005), (Ibrahim *et al*., 1997), (Domínguez & Nandi, (2008, 2009a, 2009b)), (Sheshadri & Kandaswamya, 2007), (Hassanien, 2007), (Subashini *et al*., 2010), (Song *et al*., 2009). Table 5.1 indicates the number of mammogram samples acquired from the UMMC and the MIAS dataset.

Table 5.1: Mammography data acquired from UMMC and MIAS database

Data Source	Malignant Samples	Benign Samples	Normal Samples	Database Samples
University Malaya Medical Centre (UMMC)	58	46	156	**260**
mini-MIAS Database of Mammograms (Suckling *et al*., 2004)	52	60	210	**322**
Total Samples	110	106	366	**582**

Both datasets in Table 5.1 consist of standard images of dense, fatty and fatty-glandular breasts, which are classified into three major categories: malignant, benign and normal. Samples of digital mammography images acquired from the UMMC and the MIAS datasets are shown in Figures 5.3 and 5.4 respectively. Based on the visual inspection of the mammography images, MCCs were found in most malignant and benign cases apart from mass lesions.

The total number of mammographic images obtained from UMMC is limited, i.e., a total of 256 mammographic images, as indicated in Table 5.1. The reason for this is due to the fact that UMMC has only recently implemented digital mammography, that is, in 2008. Thus, over a course of nearly two years from 2008 to 2010, only a limited number of malignant and benign cases are available in digital format.



(a) (b)

(c) (d)

Figure: 5.3: Mammography images acquired from UMMC
(Dept. of Biomedical Imaging at UMMC, 2010)

The UMMC and MIAS mammography images are digitized at 200 micron pixel edge, with a size of size of 1024 × 1024 pixels. Each pixel in the grayscale mammogram image represents the pixel intensity in the range of [0, 255] (8-bit).

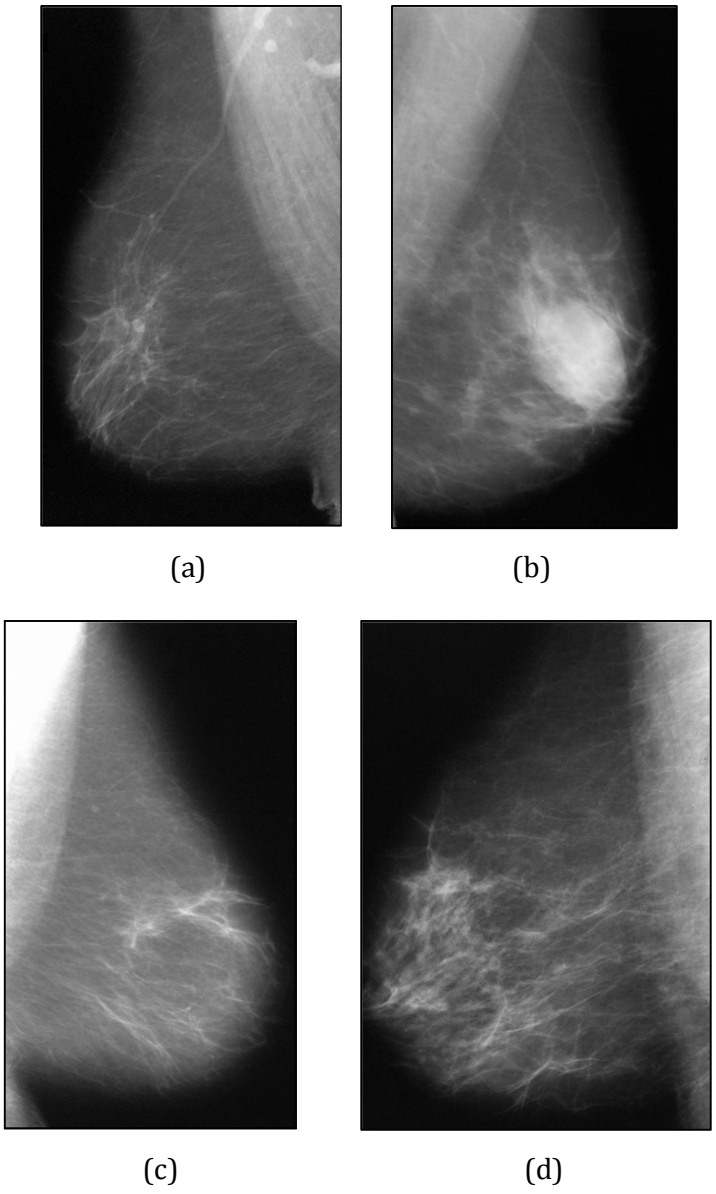


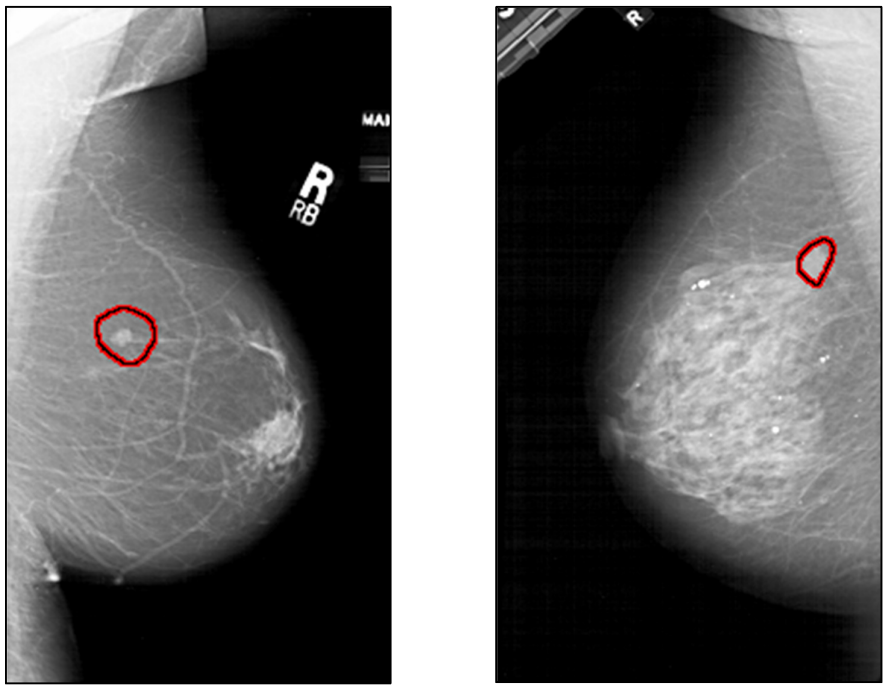
Figure: 5.4: Mammography images acquired from MIAS (Suckling *et al.*, 2004)

Other information obtained for the mammogram images from both datasets include the *Ground Truth* (GT) data and markings. The GT data contains the diagnostic results of the radiologists' interpretation of mammograms, i.e. the location of benign and malignant abnormalities, as indicated in Figure 5.5. The GT data for the UMMC dataset was provided by expert radiologists who have diagnosed and treated Malaysian patients with benign and malignant abnormalities. The GT data for the MIAS database (Suckling *et al.*, 1994) was

acquired online together with the mammography dataset. The common information retrieved from the GT data for both the UMMC and the MIAS dataset are:

1. The centre of location of the abnormality (malignant/benign) in the digital images in the format of (x, y) co-ordinates.
2. The approximate radius in pixels of a circle enclosing the abnormality area.

In this research, this GT data is identified for the purpose of extracting ROIs (abnormal regions) from the mammography images. The following sections in this chapter discuss the modeling of the proposed framework (system) as outlined in Section 5.1.



(a) (b)

Figure: 5.5: Ground Truth (GT) markings by expert radiologists on acquired mammography datasets

5.3 Mammogram Image Processing

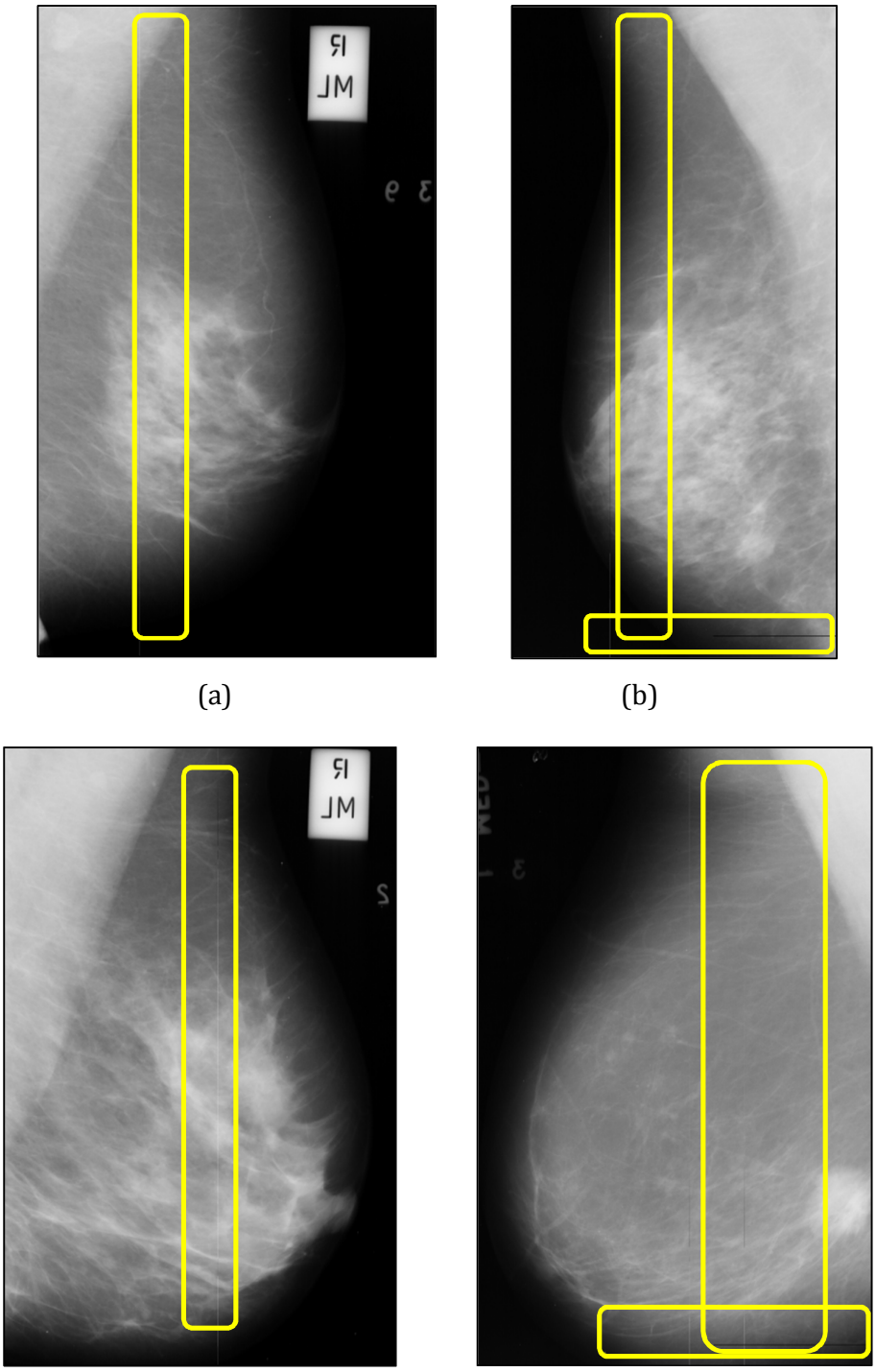
5.3.1 Image Preprocessing

Digital mammograms typically contain artifacts in the form of labels, wedges and markers in the background region. These artifacts are usually radiopaque such that that they are not transparent to radiation. The major problem with the precise segmentation of the breast region is due to the existence of such artifacts, which may cause trivial segmentation algorithms to fail. The mammogram preprocessing stage indicated in Figure 5.2 involves noise removal and radiopaque artifact suppression in order to suppress the background (black pixels) in the mammogram images. Another purpose of mammogram preprocessing is to improve the reliability and robustness of the mammogram segmentation, as discussed in the following sections.

5.3.1.1 Noise Removal

The grayscale mammography images are digitally represented using the MATLAB Image Processing Toolbox (Image Processing Toolbox, 2010). The digital images are mathematically represented using equation (3.6), where the range of intensity values of the acquired mammogram images has [0,255] gray levels.

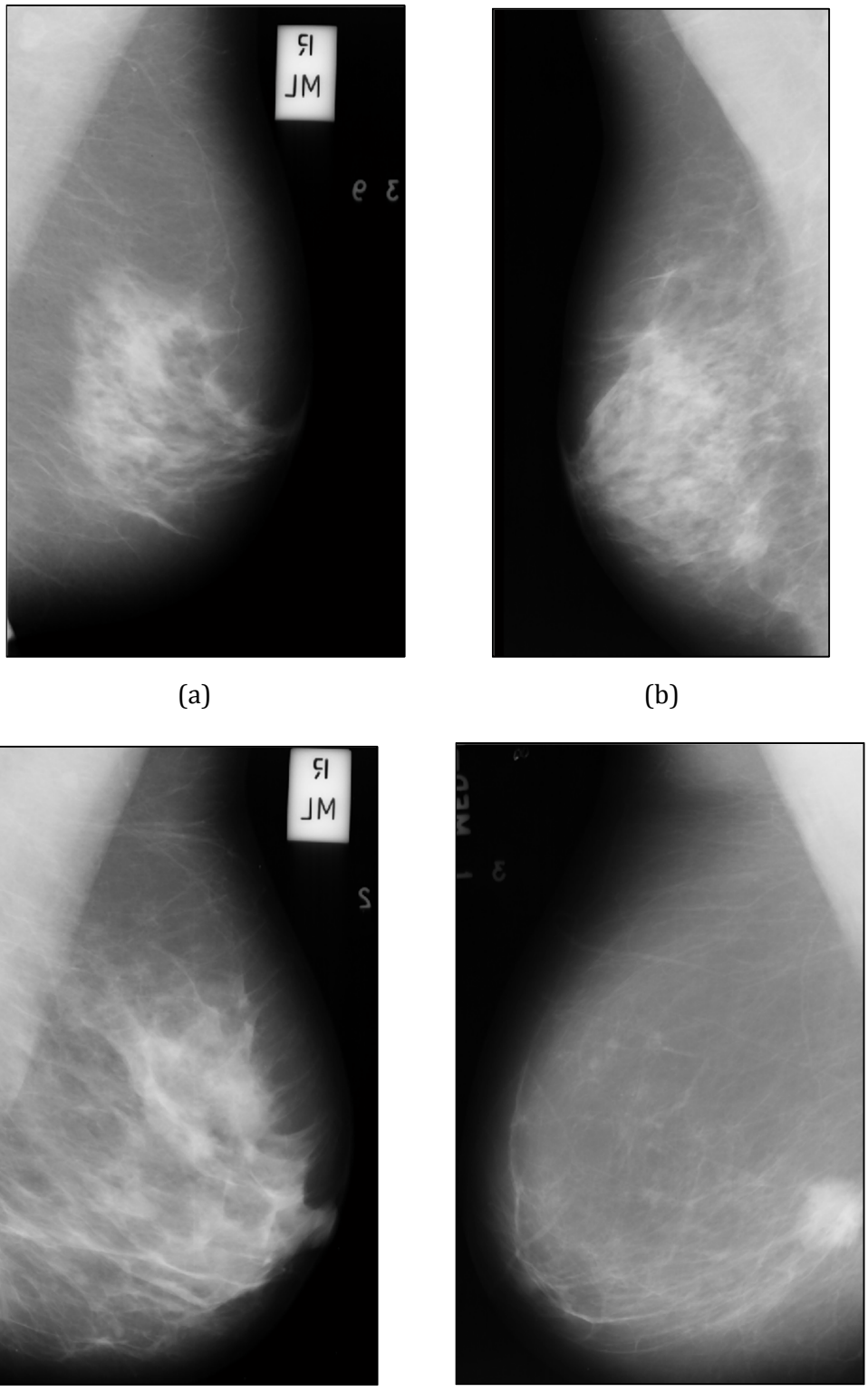
Digitization noises such as horizontal and vertical lines tend to appear on most of the mammogram images, as shown by indication markers in Figure 5.6. These noises are removed from the mammogram images by applying a *Two-dimensional Median Filtering* approach (Lim, 1990) in a 3-by-3 connected neighborhood.



(a) (b)

(c) (d)

Figure: 5.6: Digitization noises (lines) in mammographic images



(a) (b)

(c) (d)

Figure: 5.7: Mammogram images after noise removal using 2D median filtering

In the 2D median filtering approach, each output pixel contains the median value in the 3-by-3 neighborhood around the corresponding to the pixel of interest in the image. The edges of the image however, are replaced by zeros (total absence or black color). This does not affect the original image, since the ROI of the image does not include the edges or boundaries of the image. Figure 5.7 shows the mammogram images in Figure 5.6 after applying the 2D median filtering approach. The digitization noises, i.e. the horizontal and vertical lines are removed from the mammogram images without affecting the breast profile.

5.3.1.2 Radiopaque Artifact Suppression

Mammogram artifacts such as identification labels, markers, and wedges are radiopaque such that that they are not transparent to radiation. These artifacts are small emulsion continuity faults on the mammogram films which look like calcifications. Suppressing radiopaque artifacts and the background region within a mammogram image, increases the region homogeneity and also improves the reliability and robustness of the breast profile separation. A mammogram artifact suppression algorithm based on the area morphology presented by Wirth *et al.* (2004) is adopted in this research (Yapa & Harada, 2008), (Wirth *et al.,* 2007).

In order to use the area morphology approach presented by Wirth *et al.* (2004), the grayscale [0,255] mammogram image needs to be transformed into the binary [0,1] format. The simplest technique for transforming a grayscale image into binary is by using thresholding (see Section 3.4.3.1). In order to convert a grayscale image into binary, a grayscale *threshold* for that image needs to be determined in order segment the artifacts and the background, while keeping the breast-skin edges in contact, so as not to lose information from the breast profile.

In background and artifact suppression, a *global threshold* (T) is a user-determined value that is used to optimally segment the background region and the radiopaque artifacts from the breast profile for a mammogram image dataset. In order to determine a *global threshold* (T) for a mammogram image dataset, a trial and error procedure is typically used, where the segmentation performance of all the mammogram images is evaluated for all possible threshold levels. In this research, visual inspection of the segmented mammogram images determines the global threshold to be: $T = 18$, for all possible threshold levels in between 0 to 255. Figure 5.8(a) and Figure 5.8(c) represent the histograms of the mammogram images in Figure 5.9(a) and Figure 5.9(c) respectively. Figure 5.9(b) and Figure 5.9(d) show the resulting histograms obtained after thresholding the images corresponding to Figures 5.9(a) and 5.9(d) respectively. As observed from Figures 5.8 and 5.9, applying a threshold value of $T = 18$ sets all the pixels within the intensity range of 0 to 18 to a single value of 0 (background pixel).

The resulting binary images after global thresholding are indicated in Figure 5.9, where the breast profile is segmented from the background region. Next, radiopaque artifacts are suppressed by applying Morphological Operations (in Section 3.4.4) on the binary images as illustrated in Figure 5.11. The artifact suppression algorithm by Wirth *et al.* (2004) is modified in order to suit both the *local* and *external* datasets and is presented as follows:

Mammogram Artifact Suppression Algorithm

1. All objects present in the binary image are labeled. The binary objects consist of the radiopaque artifacts and the breast profile as indicated in Figure 5.11(b).

2. The ‘area’ (actual number of pixels in a region) is calculated for all objects (regions) in the binary image as shown in Figure 5.11(b).

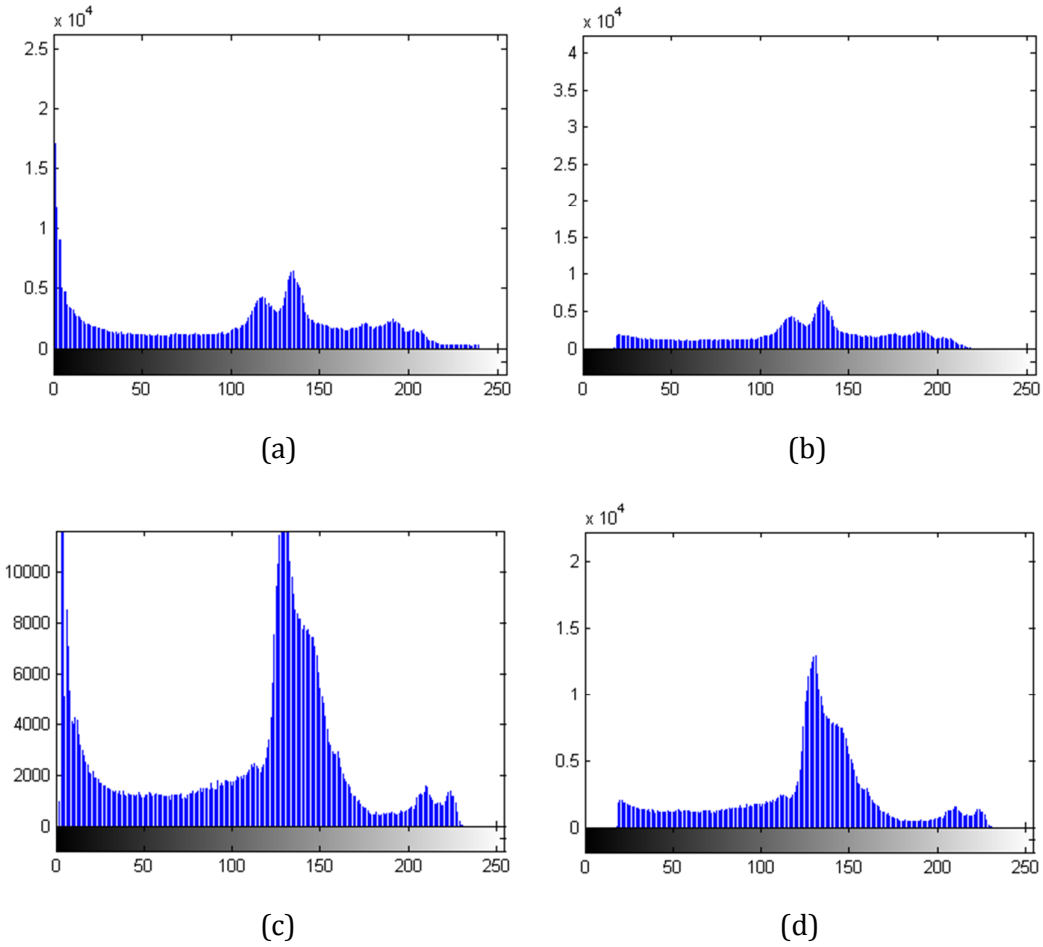


Figure: 5.8: Histograms after applying global thresholding using $T = 18$.
(a) Original histogram of mammogram image in Figure 5.7(a)
(b) Histogram of mammogram image in Figure 5.7(a) after thresholding.
(c) Original histogram of mammogram image in Figure 5.7(d)
(d) Histogram of mammogram image in Figure 5.7(d) after thresholding.

3. The largest object (area calculated in Step(2)) in the binary image in Figure 5.11(b) is then selected. This operation opens a binary image and removes all objects in the binary image, retaining the largest object, which is the breast profile region as shown in Figure 5.11(c). This operation uses an 8-connected neighborhood for a two-dimensional connective.

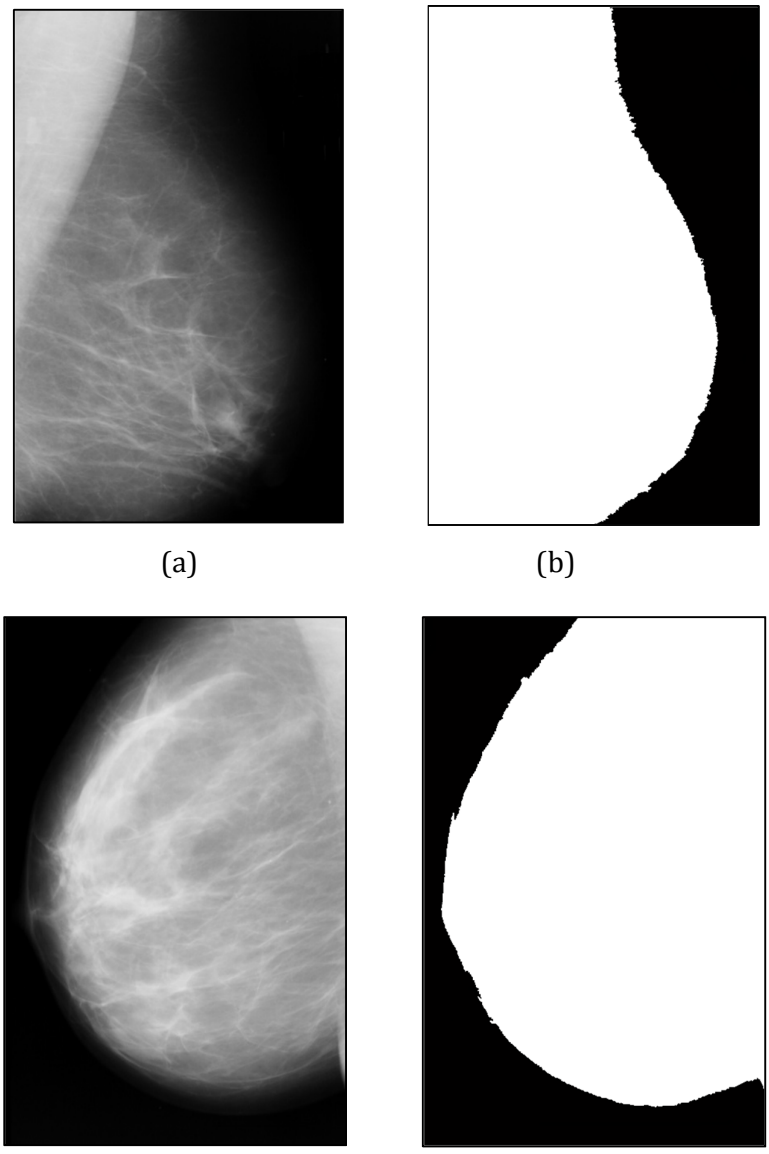


Figure: 5.9: Thresholding for segmentation of the breast profile region from the background region. (a) Original mammogram image
(b) Mammogram image in Figure 5.9(a) after breast profile separation
(c) Original mammogram image
(d) Mammogram image in Figure 5.9(c) after breast profile separation

4. Next, an operation to remove isolated pixels (individual 1's that are surrounded by 0's) and reduce distortion is applied the binary image. This algorithm checks all pixels in the binary image and if the 8-8 connected

neighborhood around the pixel of interest (pixel with value 1) consists of all 0's (isolated pixel), then the pixel of interest is set as 0 (background). An example of an isolated pixel is given by matrix X:

$$X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{5.1}$$

5. Another operation is applied to the binary image to smoothen less visible noise. This simple algorithm checks all pixels in the binary image and sets a pixel to 1 if five or more pixels in its 3-by-3 neighborhood are 1's, otherwise, it sets the pixel to 0.

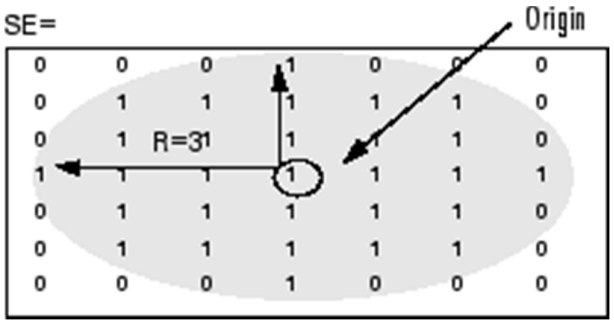
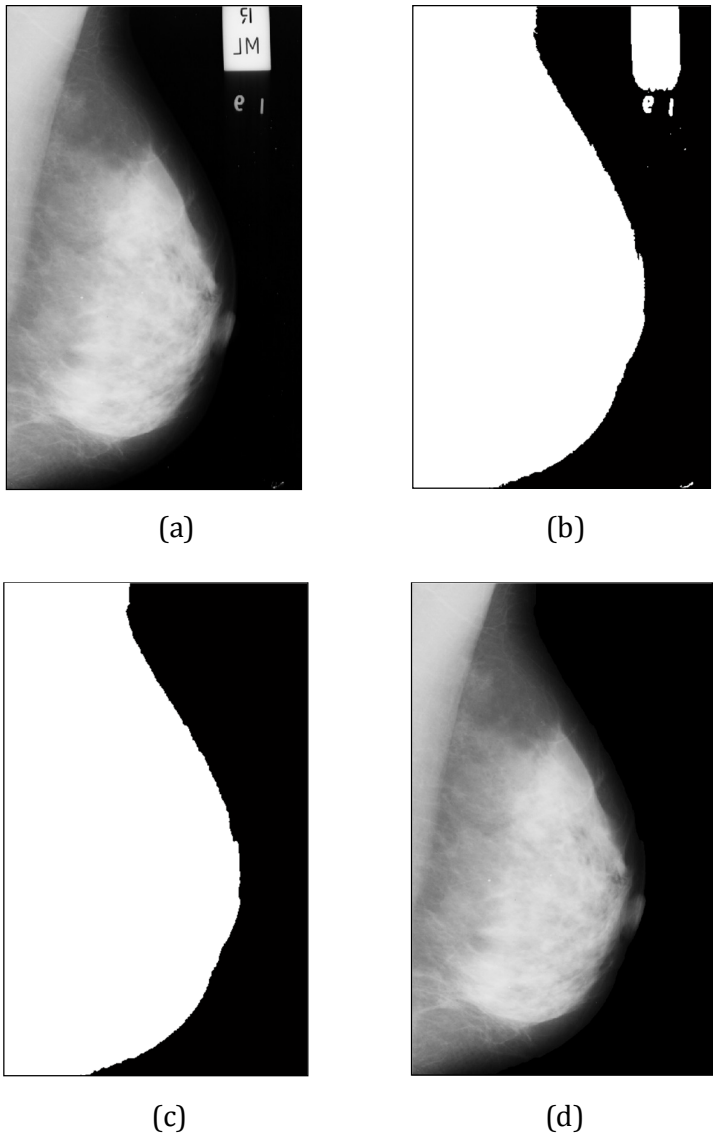


Figure: 5.10: Flat disk-shaped morphological structuring element (STREL)

6. The binary image is *eroded* (see Section 3.4.4.2) using a flat, disk-shaped morphological structuring element (STREL) as shown in Figure 5.10. Using a trial and error procedure, the optimum value of the radius (R) for the acquired mammographic image dataset is found to be $R = 3$.

7. Next, the binary image is *dilated* (Section 3.4.4.1), using the same STREL object adopted in Step (6).

8. The holes in the binary image are filled using a simple algorithm to fill all holes in the binary image, where a *hole* is defined as a set of background pixels that cannot be reached by filling in the background from the edge of the image.



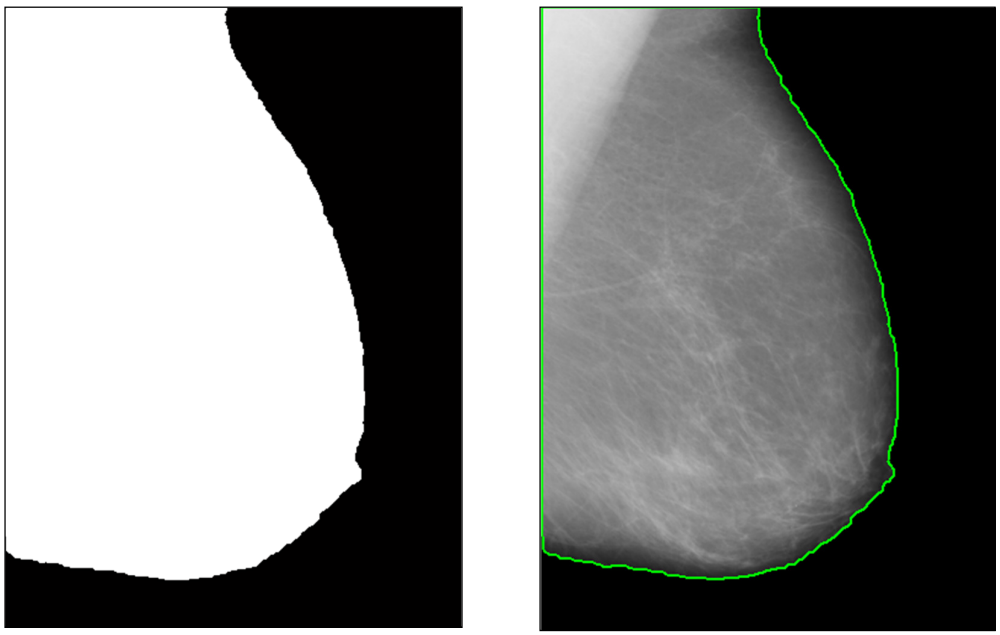
(a) (b)

(c) (d)

Figure 5.11: Suppression of radiopaque artifacts

(a) Original grayscale image with artifact and label
(b) Thresholded image using a value of T = 18 ($T_{norm} = 0.0706$)
(c) Selection of the largest object with respect to Area
(d) Grayscale image with radiopaque artifacts suppressed

9. The binary image obtained in Step (8), (Figure 5.11(c)), is multiplied with the original mammogram image in Figure 5.11(a) to obtain Figure 5.11(d). To illustrate this process, consider the binary image obtained from Step (8) (see Figure 5.11(c)) as x and the original mammogram image as y (see Figure 5.11(a)). Then each element in array x is multiplied with the corresponding element in array y to return the result in the corresponding element of the output array z, which is the grayscale image with the background and the radiopaque artifacts suppressed, as shown in Figures 5.11(d) and 5.12(b).



(a) (b)

Figure 5.12: Segmented breast profile region
(a) Binary image of thresholded mammogram
(b) Grayscale image after background and artifact suppression

The grayscale image obtained in Step (9) as indicated in Figure 5.12(b), has radiopaque artifacts suppressed and the breast profile region separated from the

background. The grayscale images resulting from this process are further used to perform mammogram segmentation as indicated in the following section.

5.3.2 Image Segmentation

The presence of the pectoral muscle in mammogram images effects the results of intensity based image processing methods and can bias procedures in the detection of masses and MCCs, as discussed in Section 3.3.2. Several research groups indicated that during the computerized detection of masses and MCCs the pectoral muscle should be excluded from processing (Kwok *et al*. 2004), (Raba *et al*. 2005), (Nicolaou *et al*. 2008), (Xu *et al*. 2007), (Ferrari *et al*. 2004), (Mirzaalian *et al*. 2007), (Bajger *et al*. 2005). In order to confirm these findings, the GT data of the mammograms acquired in this research was analyzed, which indicated that no mass lesions or MCCs appear in the pectoral muscle region. This justifies the fact to remove the pectoral muscle from the mammogram images.

The *Seeded Region Growing* (SRG) technique (in Section 3.4.3.3.2) as described in the framework in Figure 5.1, is proposed in this research as a suitable technique for the purpose of pectoral muscle segmentation. In this research, the pectoral muscle segmentation is implemented in MATLAB, which consists of the following four stages:

Stage 1 — Determining Orientation of the Breast Profile

The breast orientation (left-side or right-side) in each mammogram image needs to be determined prior to performing Seeded Region Growing (SRG) for pectoral muscle segmentation. In order to perform SRG, a seed needs to be placed inside the pectoral muscle, hence determining the breast orientation is crucial.

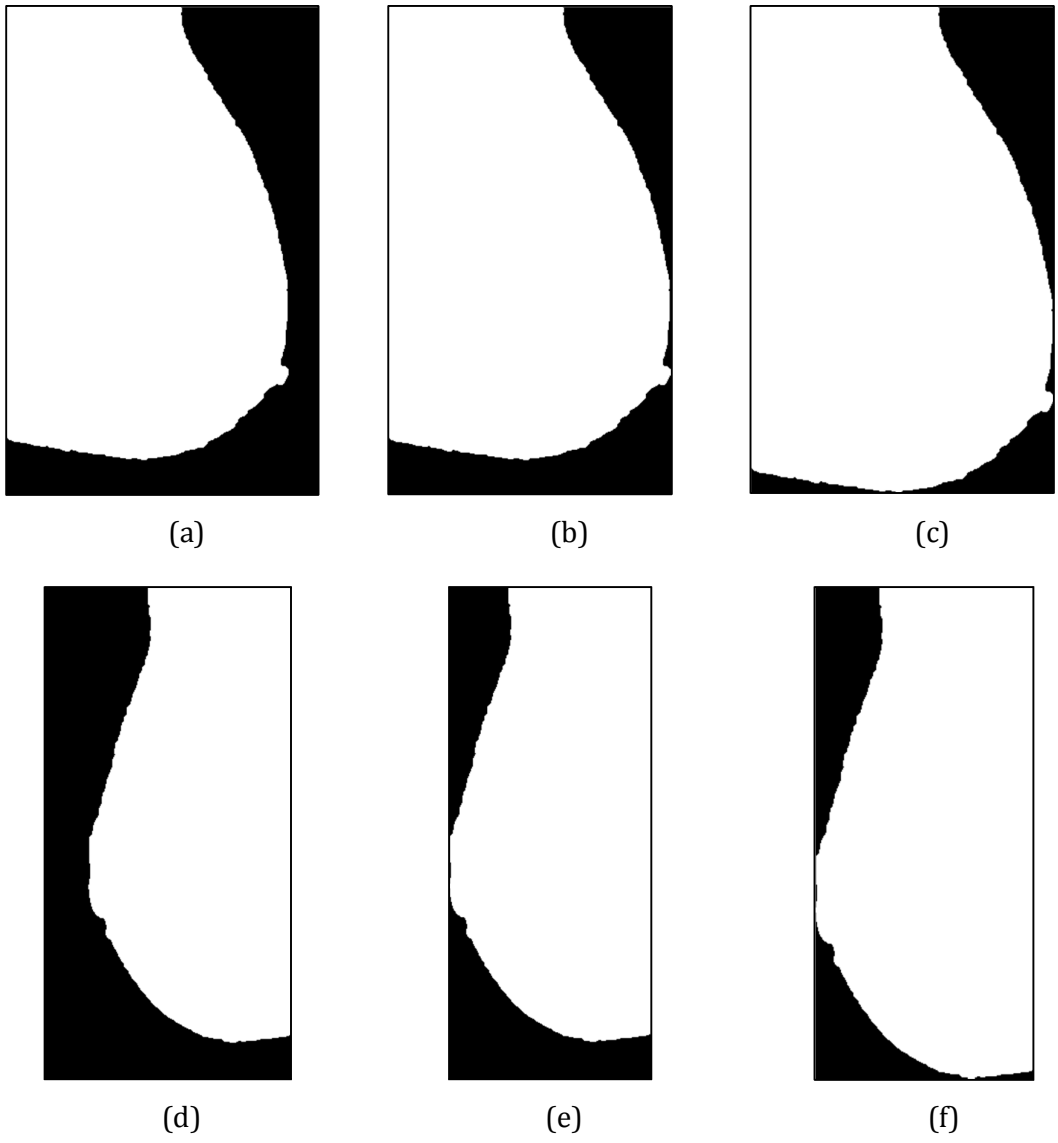


Figure 5.13: Cropping breast profile in mammogram images to image borders
(a) Binary image with right oriented breast
(b) Binary image in Figure 5.13(a) cropped from the left and right
(c) Binary image in Figure 5.13(a) cropped from the top and bottom
(d) Binary image with left oriented breast
(b) Binary image in Figure 5.13(d) cropped from the left and right
(e) Binary image in Figure 5.13(d) cropped from the top and bottom

When the pectoral muscle is located on the top left corner (see Figure 5.7(a)) of the mammogram the breast points towards the right side, hence the breast is termed as right oriented. Similarly, when the pectoral muscle is located on the top right

corner (Figure 5.7(b)) the breast points towards the left side, hence the breast is left oriented. A simple algorithm for determining the breast profile orientation is developed in this research, which is as follows:

(a) In order to determine the breast profile orientation (left-side or right-side) of mammograms using an automated procedure, the binary image in Figure 5.12(a) is used. As indicated in Figure 5.13(b) the binary image is cropped left to right and top to bottom, such that the breast profile object touches all four borders (left, right, top and bottom) of the binary image, as shown in Figure 5.13(c).

(b) After the breast profile is cropped, the sum of the pixel intensities in *first* and *last* five columns of the binary images in Figure 5.13(c) and Figure 5.13(f) is calculated. To determine the breast profile orientation, intensities of the pixels in the shaded area of equations (5.2) and (5.3) are calculated, which represent the pixels in first and last five columns of the binary image respectively.

$$f = \begin{bmatrix} f(0,1) & f(0,2) & f(0,3) & f(0,4) & f(0,5) & \dots & f(0,N) \\ f(1,1) & f(1,2) & f(1,3) & f(1,4) & f(1,5) & \dots & f(1,N) \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ f(M,1) & f(M,2) & f(M,3) & f(M,4) & \boldsymbol{f(M,5)} & \dots & f(M,N) \end{bmatrix} \tag{5.2}$$

$$f = \begin{bmatrix} f(0,1) & \dots & f(0,N-4) & f(0,N-3) & f(0,N-2) & f(0,N-1) & f(0,N) \\ f(1,1) & \dots & f(1,N-4) & f(1,N-3) & f(1,N-2) & f(1,N-1) & f(1,N) \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ f(M,1) & \dots & \boldsymbol{f(M,N-4)} & f(M,N-3) & f(M,N-2) & f(M,N-1) & f(M,N) \end{bmatrix} \tag{5.3}$$

In equations (5.2) and (5.3), the binary images have M rows and N columns, or in other words, the images are of $M \times N$ dimensions. Each element of the matrix $f(M,N)$ in equations (5.2) and (5.3) corresponds to an *image pixel*, represented by a binary value [0,1]. Using the shaded area in equation (5.2) and equation (5.3), the sum of the first and last 5 columns of the binary images in Figures 5.13(c) and 5.13(f) is calculated respectively, using the following expressions:

$$sum_{first} = \left\{ \begin{array}{c} \text{sum}\big(f(0,1), f(1,1), \dots, f(M,1)\big) \\ + \text{sum}\big(f(0,2), f(1,2), \dots, f(M,2)\big) \\ + \text{sum}\big(f(0,3), f(1,3), \dots, f(M,3)\big) \\ + \text{sum}\big(f(0,4), f(1,4), \dots, f(M,4)\big) \\ + \text{sum}\big(f(0,5), f(1,5), \dots, f(M,5)\big) \end{array} \right\} \tag{5.4}$$

$$sum_{last} = \left\{ \begin{array}{c} \text{sum}\big(f(0,N-4), f(1,N-4), \dots, f(M,N-4)\big) \\ + \text{sum}\big(f(0,N-3), f(1,N-3), \dots, f(M,N-3)\big) \\ + s\text{um}\big(f(0,N-2), f(1,N-2), \dots, f(M,N-2)\big) \\ + \text{sum}\big(f(0,N-1), f(1,N-1), \dots, f(M,N-1)\big) \\ + \text{sum}\big(f(0,N), f(1,N), \dots, f(M,N)\big) \end{array} \right\} \tag{5.5}$$

where sum_{first} and sum_{last} are the sums (as integer values) of the binary values [0,1] in the first and last five columns of the binary image in Figure 5.13(c) and Figure 5.13(f).

(c) Next, the sums of the binary values (sum_{first} and sum_{last}) are used to construct a simple IF-THEN rule for identifying the breast profile as left-sided or right-sided. The decision making rule developed in this research is as given below.

Pseudo-code of for determining the breast orientation

```
if (sum_first > sum_last)
   %Breast is oriented towards the right
   %See Figure 5.13(c)
elseif (sum_first < sum_last)
   %Breast is oriented towards the left
   %See Figure 5.13(f)
end if
```

The IF-THEN logic in the pseudo-code above compares the sum of the intensity values in the binary image, namely sum_{first} and sum_{last} and accurately identifies the orientation of the breast profile of all mammogram images acquired (in Table 5.1). Using this rule, Figure 5.13(c) is identified as a right-orientated breast, while Figure 5.13(f) is identified as a left-orientated breast.

Stage 2 — Breast Profile Contrast Enhancement

Prior to pectoral muscle segmentation, the contrast of the breast profile, particularly the pectoral muscle region, needs to be enhanced. Since the pectoral muscle contains the majority of the brightest pixels in the breast profile, hence the contrast of the breast profile needs to be enhanced such that the grayscale pixels in the pectoral muscle become brighter. By using contrast enhanced images, the SRG algorithm can more reliably identify the pectoral muscle border from the breast profile.

In order to perform contrast enhancement, the mammogram image obtained in Figure 5.12(b) is used. Contrast enhancement is performed by finding the limits to contrast stretch an image, where the tolerance level is a scalar quantity, which

saturates fractions at low and high pixel levels. In this research, the default tolerance value of $t = [min\ max]$ is used, where min is the smallest grayscale value in the mammogram image and max is the highest grayscale value in the mammogram image.

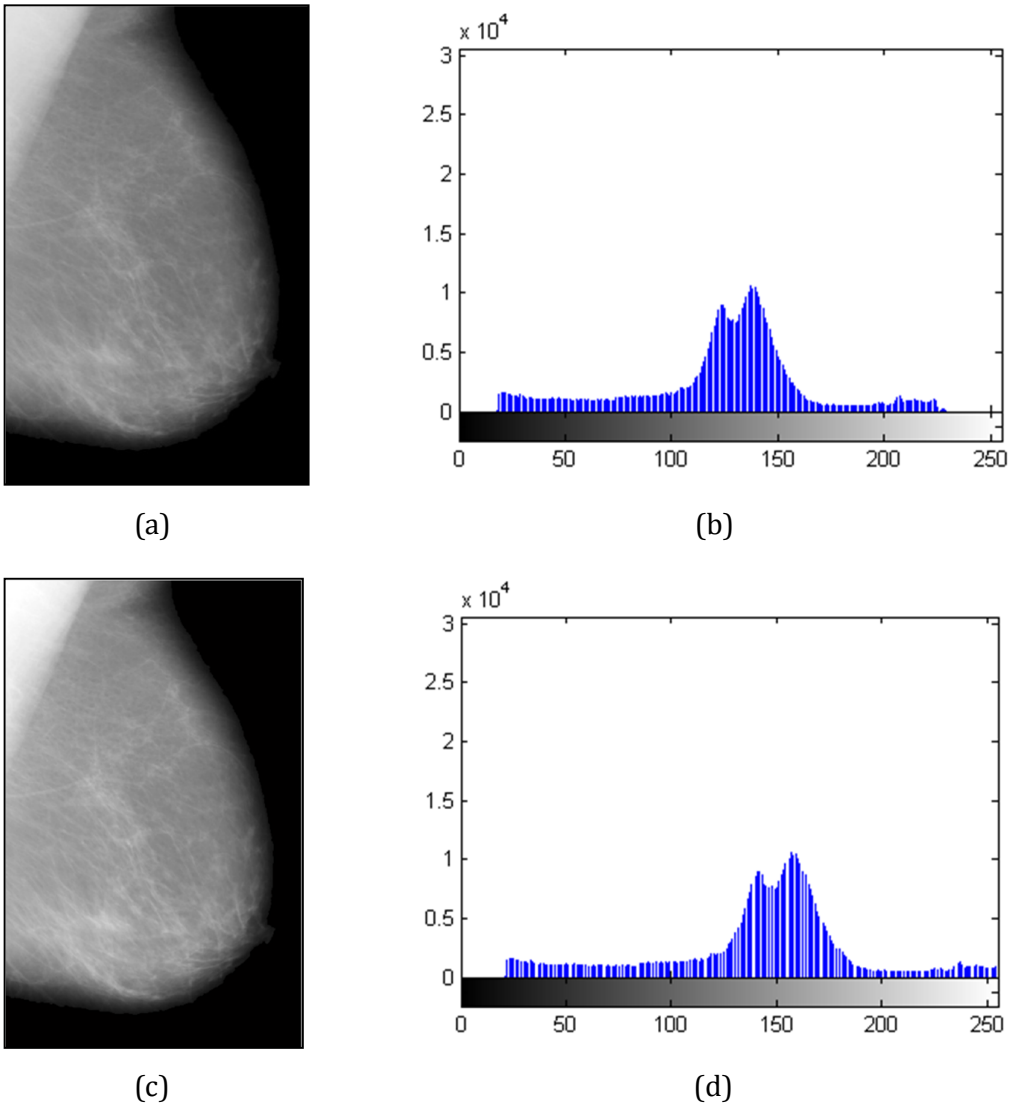


Figure 5.14: Contrast enhancement of a mammogram image

(a) Mammogram image obtained after mammogram preprocessing (Figure 5.12).

(b) Histogram of the original image in (a)

(c) Contrast enhancement applied to the mammogram image in (a)

(d) Histogram of contrast enhanced image in (c)

Figure 5.14 shows the contrast enhancement technique applied to the mammogram images. As indicated in Figure 5.14(d) the histogram of the original image in Figure 5.14(b) is stretched, which increases the number of brighter pixels.

Stage 3 — Seeded Region Growing

After the breast orientation is determined in the first stage and the breast profile contrast is enhanced in the second stage, the Seeded Region Growing (SRG) technique (Adams & Bischof, 1994) is then applied to segment the pectoral muscle as discussed in Section 3.4.3.3.2. In order to implement SRG a *seed* needs to be placed inside the pectoral muscle of the grayscale mammogram image.

Using results obtained from the first and second stages, if the breast profile is right-orientated a seed is placed in the top-left corner (pectoral muscle) of the image shown in Figure 5.15(a), corresponding to pixel $f(M,5)$ circled in equation (5.2). Similarly, if the breast profile is right-orientated, a seed is placed in the top-right corner (pectoral muscle) of the mammogram image shown in Figure 5.15(d), corresponding to pixel $f(M,N-4)$ circled in equation (5.3)). The theoretical concepts of the SRG algorithm have been presented in detail in Section 3.4.3.3.2 of this thesis. The following four steps are applied in the SRG process:

After the breast orientation is determined in the first stage and the breast profile contrast is enhanced in the second stage, the Seeded Region Growing (SRG) technique (Adams & Bischof, 1994) is next applied as discussed in Section 3.4.3.3.2 for segmenting the pectoral muscle. In order to implement SRG a *seed* needs to be placed inside the pectoral muscle of the grayscale mammogram image. Using results obtained from the first and second stages, if the breast profile is right-

orientated a seed is placed in the top-right corner (pectoral muscle) of the image shown in Figure 5.15(a), else if the breast profile is right-orientated, a seed is placed in the top-left corner (pectoral muscle) of the image shown in Figure 5.15(d). Theoretical concepts of the SRG algorithm are presented in detail in Section 3.4.3.3.2 of this thesis. The following four steps are applied in the SRG process:

a. The region is iteratively grown by comparing all unallocated neighboring pixels to the region.
b. The difference between the pixel of interests' intensity value and the region's mean is used as a measure of similarity.
c. The pixel with the smallest difference measure is allocated to the respective region.
d. The process stops when the intensity difference between the region mean and the new pixel become larger than the threshold value (maximum intensity distance).

Using a trial and error procedure for inspecting the segmentation performance of all the mammogram images for all possible threshold levels in between 0 to 255, the optimum *SRG threshold* satisfying all mammogram images (in Table 5.1) is determined to be $S = 18$. The SRG threshold is generally referred to as the maximum intensity distance satisfying all mammogram images to reliably segment the pectoral muscle from the breast profile region. The results of the segmented pectoral muscle obtained after the SRG process, is a binary image as shown in Figure 5.15(c) and Figure 5.15(f).

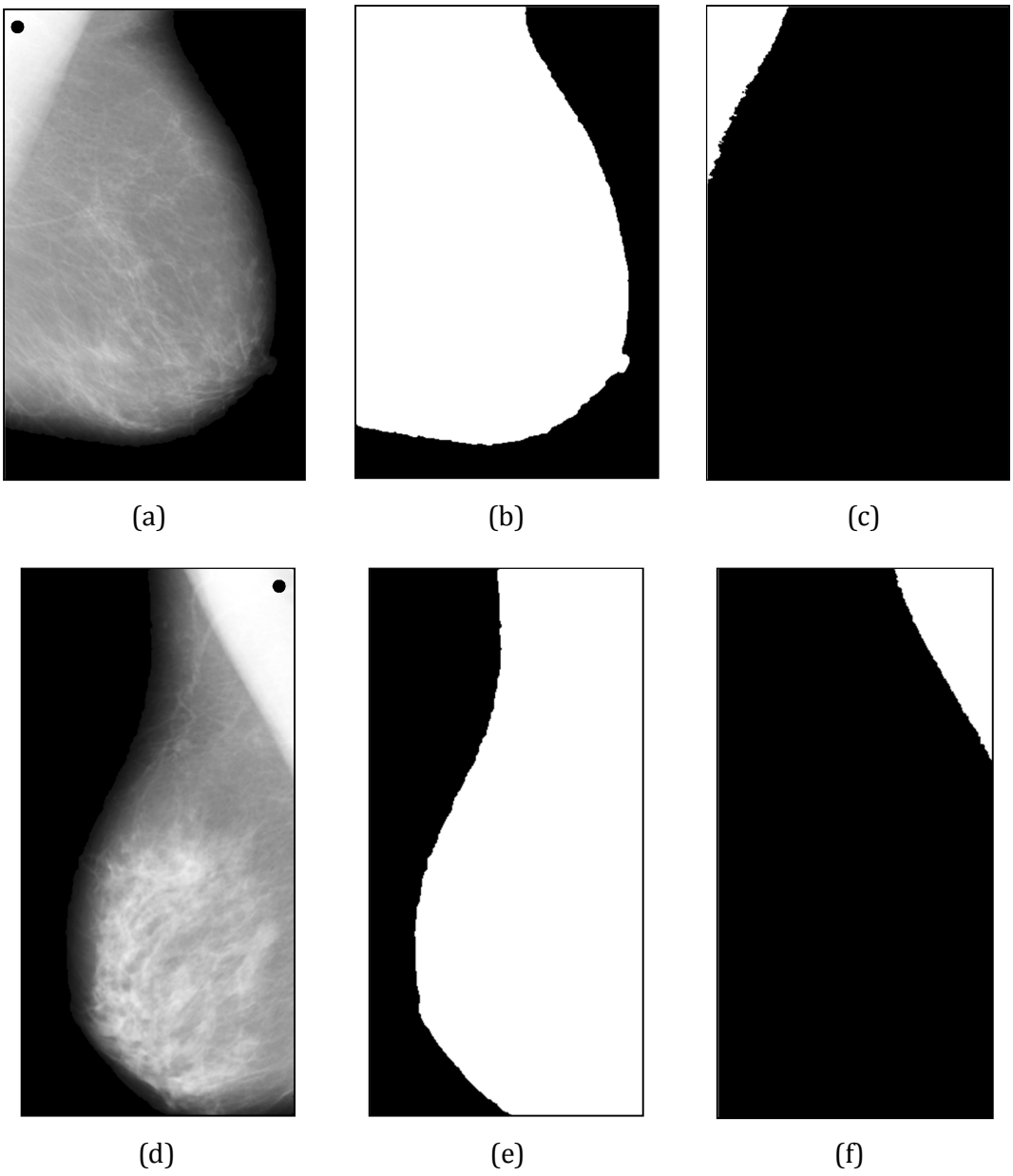


Figure 5.15: Segmentation of pectoral muscle using Seeded Region Growing

(a) Contrast enhanced breast profile right-orientated

(b) Binary image of Figure 5.15(a) showing separated breast profile

(c) Segmented pectoral muscle of Figure 5.15(a) using SRG

(d) Contrast enhanced breast profile left-orientated

(e) Binary image of Figure 5.15(d) showing separated breast profile

(f) Segmented pectoral muscle of Figure 5.15(d) using SRG

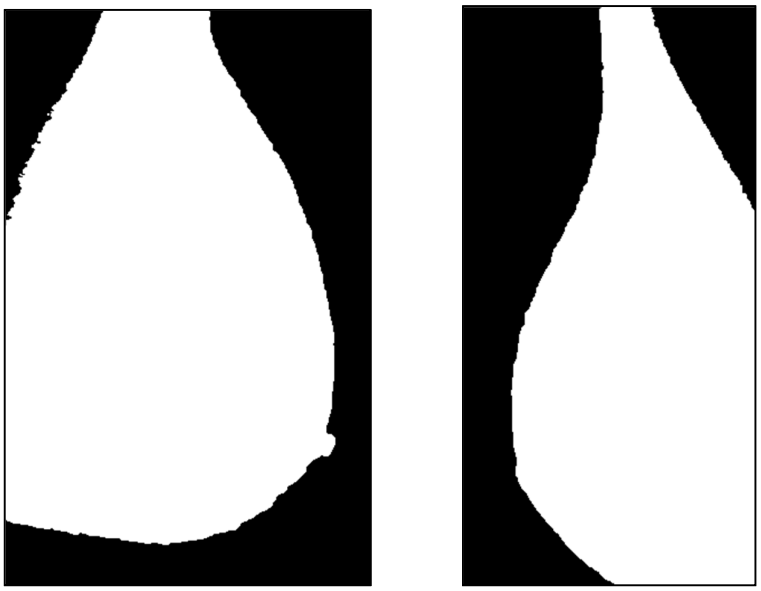
Next, the binary object in Figure 5.15(c) and Figure 5.15(f) is subtracted from the breast profile region in Figure 5.15(b) and Figure 5.15(e) respectively. To illustrate this concept, consider the breast profile in the binary image as x and the segmented pectoral muscle in the binary image as y. Then each element in array y is subtracted from the corresponding element in array x returning the difference in the corresponding element of the output array z, which is a binary image of the breast profile with the pectoral muscle segmented, as indicated in Figures 5.16(a) and 5.16(b). This process can be implemented in MATLAB by using the available built in library such as the *imsubtract* function.

Typically there is only one binary object present in the binary image (see Figure 5.16(a) and Figure 5.16(b)) after the segmentation of the pectoral muscle. However in a few cases, other smaller objects may be present, i.e. parts of the pectoral muscle near the breast profile and pectoral muscle border are retained. In order to cater for these small binary objects, which represent the segmentation noise, the following three steps are applied:

a. All objects in the binary image in Figures 5.16(a) and 5.16(b) are labeled.

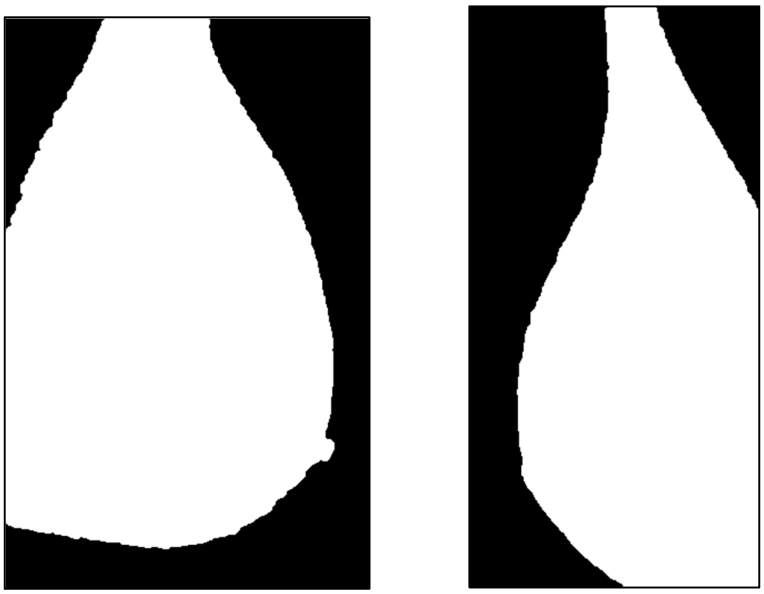
b. The 'area' (actual number of pixels in a region) is calculated for all objects (regions) in the binary images (in Figure 5.16(a) and Figure 5.16(b)).

c. The largest object (area calculated in Step(b)) in the binary images in Figure 5.16(a) and Figure 5.16(b) is selected. This operation opens a binary image and removes all objects in the binary image, retaining the largest object, which is the pectoral muscle as shown in Figures 5.15(c) and 5.15(f). This operation uses an 8-connected neighborhood.



(a) (b)

Figure 5.16: Suppression of pectoral muscle from breast profile region
(a) Subtraction of binary image Figure 5.15(c) from binary image Figure 5.15(b)
(b) Subtraction of binary image Figure 5.15(e) from binary image Figure 5.15(f)



(a) (b)

Figure 5.17: Binary images of the segmented breast profile
(a) Right-oriented breast profile after morphological operations
(b) Left-oriented breast profile after morphological operations

Next, the morphological operations from Step (4) through Step (8) of Section 5.3.1.2 are performed on the binary images in Figure 5.16 to reduce distortion and noise, fill holes and remove small objects. The resulting binary image after performing the morphological operations is shown in Figure 5.17. Comparing Figure 5.16 with Figure 5.17, it is observed that after applying morphological operations to the binary image, the boundary between the breast profile region (object) and the segmented pectoral muscle becomes smoother.

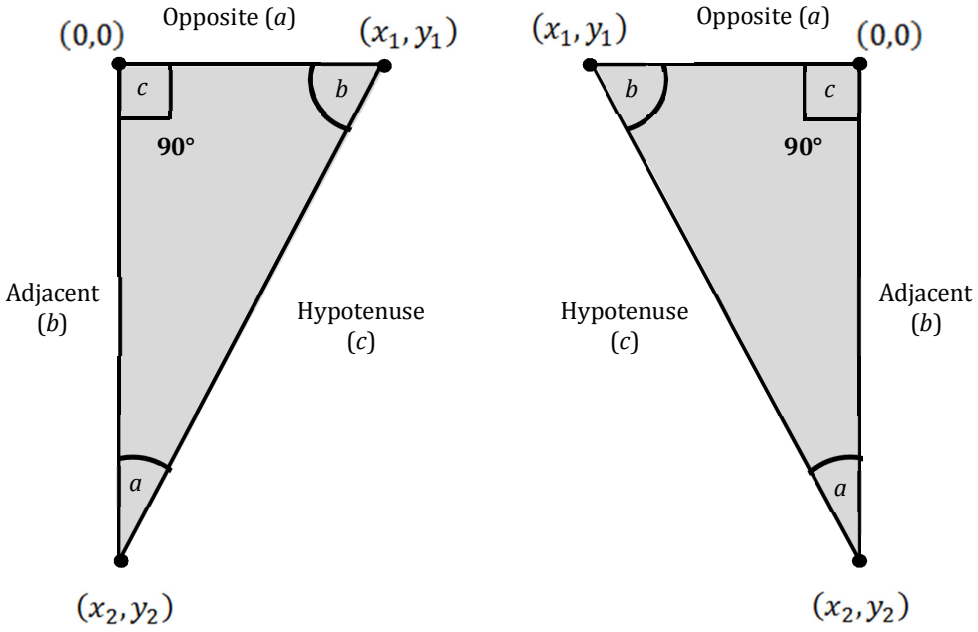
Stage 4 — Smoothening Segmented Pectoral Muscle Boundary

Since the boundary between the breast profile region and the segmented pectoral muscle contains rough (unsmooth) edges as indicated in Figure 5.17, thus a simple and effective technique is applied to smooth the boundary of the segmented pectoral muscle border. The pectoral muscle can be viewed as a right-angle triangle in Figure 5.15(c) and Figure 5.15(f), which is illustrated in Figure 5.18(a) and Figure 5.18(b) respectively.

In order to smooth the pectoral muscle boundary, i.e., the hypotenuse in Figure 5.18, the two points (x_1, y_1) and (x_2, y_2) in Figure 5.18 need to be determined in Figure 5.17, with respect to the origin (0,0). The two points (x_1, y_1) and (x_2, y_2) in the format (row, col) in the digital images are determined by iterating through the columns and rows of the binary image expressed in equation (3.6). The algorithm proposed for the detection of the two points (x_1, y_1) and (x_2, y_2) for right and left sided breasts is as follows.

Right-Oriented Breast

(a) To determine the point (x_1, y_1) in Figure 5.17(a), the first row of the binary image $\big(f(0,1), f(0,2), \dots, f(0,N)\big)$ in equation (5.2) of Figure 5.17(a) is checked for the occurrence of a 1, since 0 represents black (background) pixels. The first occurrence of 1 in the row is selected as the point (x_1, y_1) in Figure 5.18(a).



(a) (b)

Figure 5.18: Pectoral muscle viewed as a right angle triangle
(a) Pectoral muscle in Figure 5.15(c) viewed as a right angle triangle for a right-oriented breast profile
(b) Pectoral muscle in Figure 5.15(f) viewed as a right angle triangle for a left-oriented breast profile

(b) Similarly, in order to find the point (x_2, y_2) in Figure 5.17(a), the first column of the binary image $\big(f(0,1), f(1,1), \dots, f(M,1)\big)$ in equation (5.2) of

Figure 5.17(a) is checked for the occurrence of a 1. The first occurrence of 1 in the column is selected as the point (x_2, y_2) in Figure 5.18(a).

Left-Oriented Breast

a. To find the point (x_1, y_1) in Figure 5.17(b), the first row of the binary image $\big(f(0,N), f(0,N-1), \dots, f(0,2), f(0,1)\big)$ in equation (5.3) of Figure 5.17(b) is checked in reverse order for the occurrence of a 1, since 0 represents black (background) pixels. The first occurrence of 1 in the row is selected as the point (x_1, y_1) in Figure 5.18(b).

b. Similarly, in order to find the point (x_2, y_2) in Figure 5.17(b), the last column of the binary image $\big(f(0,N), f(1,N), \dots, f(M,N)\big)$ in equation (5.3) of Figure 5.17(b) is checked for the occurrence of a 1. The first occurrence of 1 in the column is selected as the point (x_2, y_2) in Figure 5.18(b).

After determining the two points (x_1, y_1) and (x_2, y_2) for the right angle triangle in Figure 5.18, an offset value r is specified in the two points, where r is an integer value in the range of $0 < r < 30$. For right-sided breasts, specifying the offset value r, transforms the two points (x_1, y_1) and (x_2, y_2) into $(x_1 - r, y_1)$ and $(x_2, y_2 + r)$. Similarly for left-sided breasts, specifying the offset value r, transforms the two points (x_1, y_1) and (x_2, y_2) into $(x_1 + r, y_1)$ and $(x_2, y_2 + r)$. Next, the *straight line equation* is used to represent the hypotenuse of the right angle triangle in Figure 5.18, which is the boundary between the breast profile region and the segmented pectoral muscle. A straight line is represented by the following expression:

$$y = mx + c \tag{5.6}$$

where m is the slope (gradient) determining the steepness of the line, and c is the intercept where the line crosses the y-axis. Using equation (5.6), the standard equation of a straight line, the value of the slope m is calculated using the following expression:

$$m = (y_1 - x_1)/(y_2 - x_2) \tag{5.7}$$

where the c-intercept is calculated using the following expression:

$$c = y_1 - (m * y_2) \tag{5.8}$$

The equation of the straight line (in equation 5.6) can be found using the MATLAB implementation. The following steps indicate how the equation of the straight line is determined.

1. The equation of the straight line is found by specifying a vector for x in equation (5.6), such that the vector x consists of y_2 points in the range of $x_2 < x < y_2$, where x_2 is always taken as 1. Using the vector x (number of points in y_2) with the determined values of m and c from equations (5.7) and (5.8), equation (5.6) which is the standard equation of a straight line is evaluated to form a straight line vector y, in order to represent the hypotenuse of the right angle triangle in Figure 5.18.

   Since the straight line obtained in vector y consists of floating point values, hence the values are rounded off to the nearest integer. The process of rounding off can be implemented in MATLAB by using the available built in library such as the $round$ function.

2. As the straight line (hypotenuse in Figure 5.18) needs to be superimposed on the boundary between the breast profile region and the segmented pectoral muscle in Figure 5.17, thus, vectors x and y (in equation (5.6)) calculated in Step 1 are combined together to form a data matrix of the straight line. The process of matrix concatenation can be implemented in MATLAB by using the available built in matrix operations.

   The data matrix of the straight line obtained is superimposed on Figure 5.17 in order to smooth the boundary between the breast profile region and the segmented pectoral muscle. The technique for superimposing the straight line on Figure 5.17 is different for left and right oriented breasts.

   For right-oriented breasts, the data matrix obtained in Step 2 is used with the binary image in Figure 5.17(a) in order to smooth the boundary between the breast profile region and the segmented pectoral muscle. To illustrate this process in simple terms, after the straight line is superimposed on the binary image in Figure 5.17(a) using the data matrix, all pixel values on the *left side* of the straight line are set to 0 (background pixel), which results in a smooth pectoral muscle boundary as indicated in Figure 5.19(a).

   Similarly, for left-oriented breasts, the data matrix obtained in Step 2 is used with the binary image in Figure 5.17(b) in order to smooth the boundary between the breast profile region and the segmented pectoral muscle. To illustrate this process, after the straight line is superimposed on the binary image in Figure 5.17(b) using the data matrix, all pixel values on

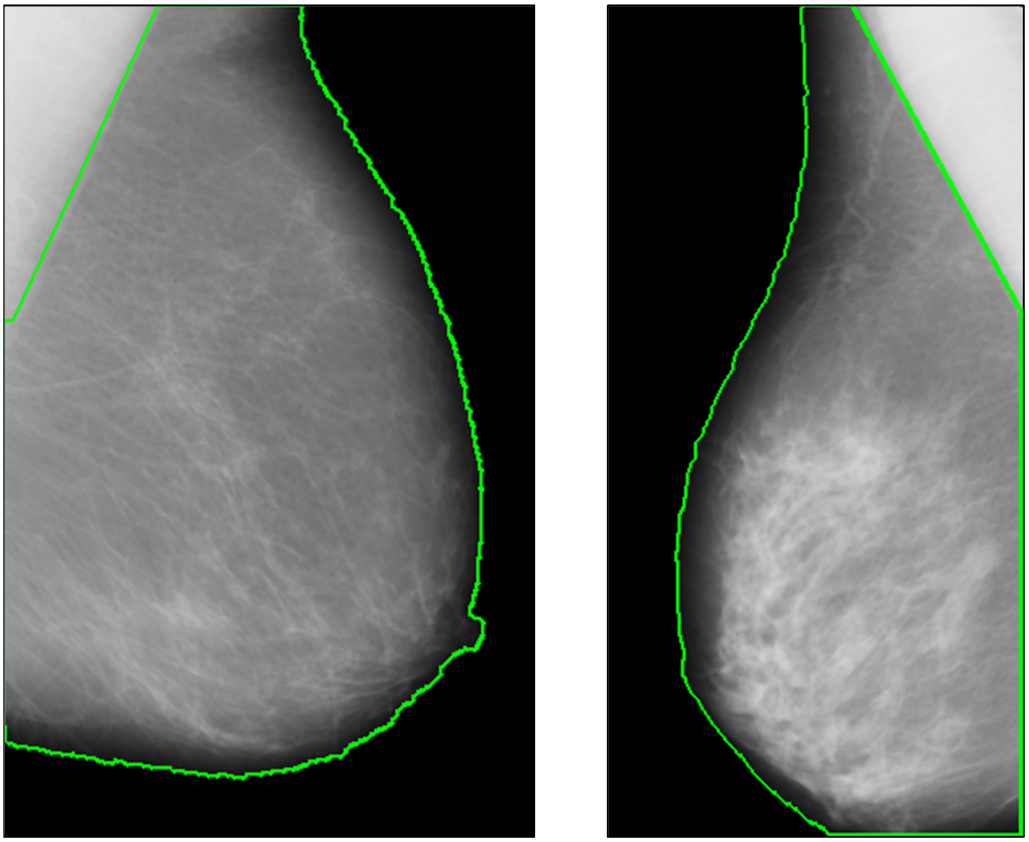
the *right side* of the straight line are set to 0 (background pixel), which results in a smooth pectoral muscle boundary as indicated in Figure 5.19(b).

(a) (b)

Figure 5.19: Binary images after pectoral muscle boundary straightening
(a) Right-oriented breast profile applying straight line
(b) Left-oriented breast profile applying straight line

The *erosion* and *dilation* operations from Step (6) and Step (7) of Section 5.3.1.2 are performed on the binary images in Figure 5.19 to remove small objects formed after superimposing the straight line. Finally, the binary image obtained in from Step 1 is multiplied with the original mammogram image in Figure 5.12(b), the result of which produces Figure 5.20.

The segmented mammograms obtained after suppressing the pectoral muscle using this technique, are indicated in Figure 5.21(a) and Figure 5.21(c), where the histograms are shown in Figure 5.21(b) and Figure 5.21(d) respectively. From the histograms in Figure 5.21(b) and Figure 5.21(d), it is observed that removing the pectoral muscle in the breast profile reduces the amount of brighter pixels in the grayscale image. This is because the majority of brighter pixels in the breast profile contribute to the pectoral muscle.



(a) (b)

Figure 5.20: Pectoral muscle segmentation from mammogram
(a) Pectoral muscle segmentation of right-oriented breast profile
(b) Pectoral muscle segmentation of left-oriented breast profile

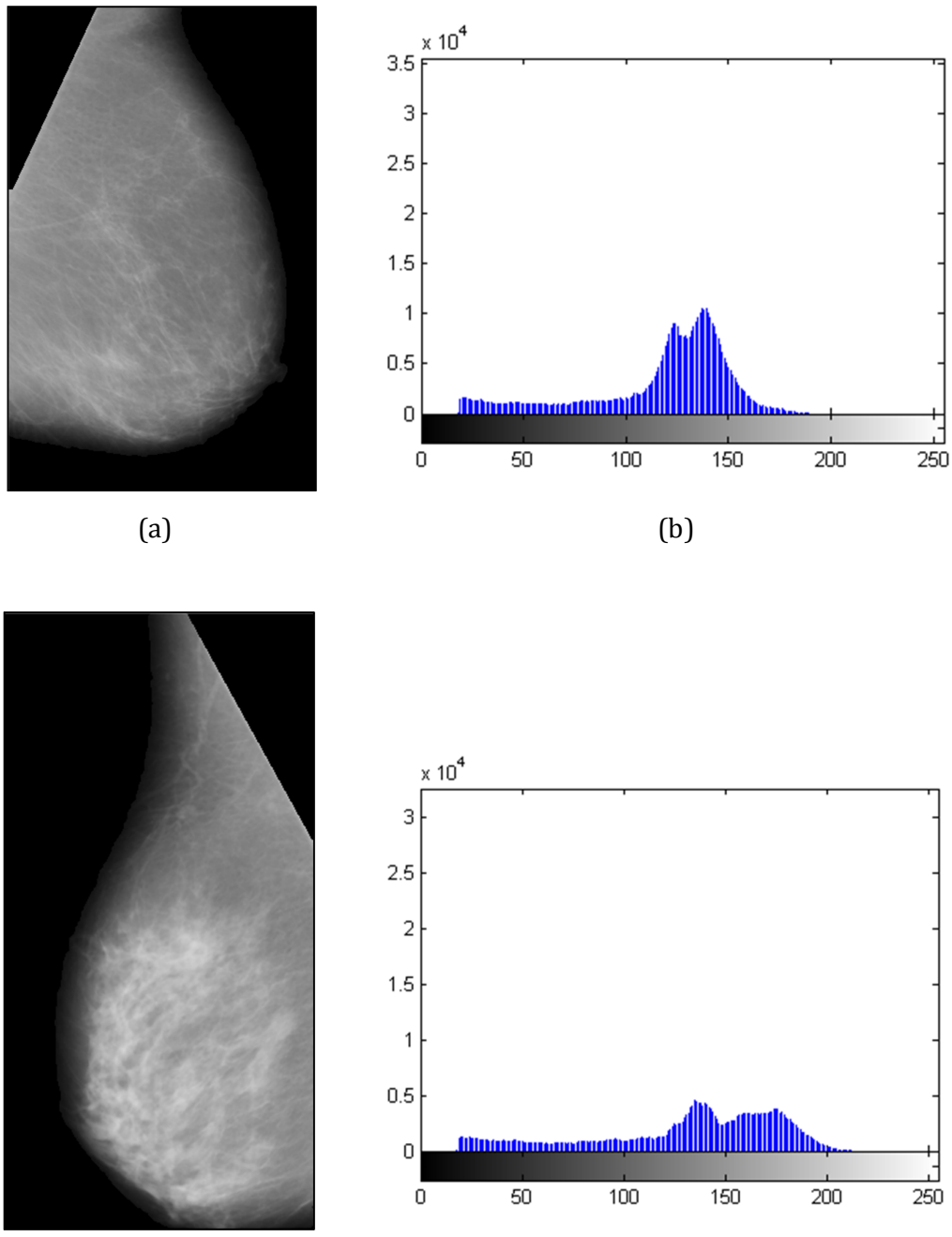


Figure 5.21: Grayscale image after pectoral muscle segmentation

(a) Grayscale image of right-orientated breast profile

(b) Image histogram of Figure 5.21(a)

(c) Grayscale image of left-orientated breast profile

(d) Image histogram of Figure 5.21(c)

5.4 Texture Feature Extraction and Selection

5.4.1 Region of Interest (ROI) Selection

Features (heuristics) cannot be directly computed from the segmented mammogram images obtained in Figure 5.21, since they will bias the detection results. So, features need to be computed only from the abnormal (malignant and benign) regions of the breast profile, while excluding all other unimportant parts of the breast tissue. In this case, the GT data (in Section 5.2.1) acquired with the mammography datasets is used in order to extract the Region of Interests (ROIs) or *abnormal regions* (malignant and benign cells) from the segmented breast profile in Figure 5.21(a) and Figure 5.21(c).

The information obtained for the malignant and benign cases from GT data and markings as presented in Section 5.2.1 is:

1. The centre of location of the abnormality (malignant/benign) in the images in (x, y) co-ordinates.
2. The approximate radius in pixels of a circle enclosing the abnormality area.

In this research, the location of the centre of the abnormality is used for the purpose of ROI extraction, where the ROI is defined as an abnormality (malignant or benign) in the mammogram. Since GLCM texture features can only be computed for image data represented in the form of a 2D matrix having a fixed shape of N rows and M columns, for this purpose, the ROI is selected in the shape of a *square*, rather than a circle as in the GT data.

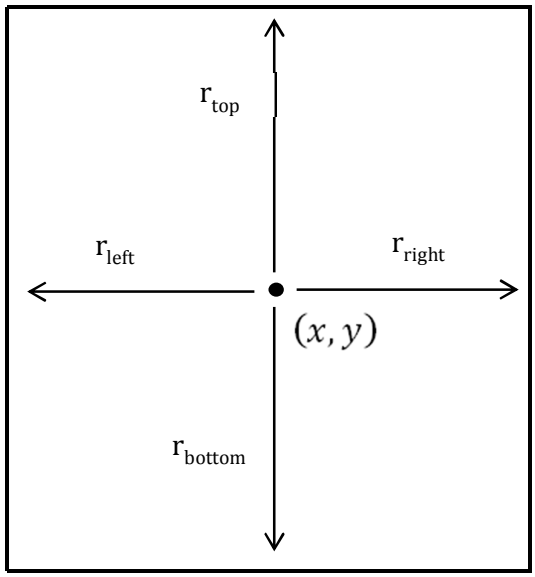


Figure 5.22: Extraction of samples using different "square" ROI sizes

The ROIs of the segmented mammogram images in Figure 5.21 are extracted using Figure 5.22 such that the centre of location of the abnormality i.e., the (x, y) co-ordinates in Figure 5.22 are the origin pixel (0,0) of the ROI in the mammogram. In this research, to estimate the size of mammographic abnormalities using a square ROI (see Figure 5.22), it is more relevant to consider the *diameter* of the abnormality. Since, the diameter is two times the radius, thus, different sizes of the diameter are used in this research, in order to determine the most optimum ROI size.

Through the analysis of the GT data, the minimum and maximum diameter in pixels of a circle enclosing all malignant and benign abnormalities is found to be 48 and 130 pixels respectively. Using this information the most common ROI sizes enclosing the majority of malignant and benign abnormalities are determined from the GT data, which are: 48 × 48 pixels, 64 × 64 pixels, 96 × 96 pixels, 110 × 110 pixels, 128 × 128 pixels, 136 × 136 pixels and 148 × 148 pixels.

The (x, y) co-ordinates of the centre of the abnormality represent the ROI origin $(0,0)$ or the ROI centre. In order to extract ROIs, the (x, y) co-ordinates from the GT data need to be represented by using integers as offset values.

As observed from Figure 5.22, the offset values representing the size of the square ROI are: r_{top}, r_{bottom}, r_{left} and r_{right}. The offsets used to extract ROIs of size 48 × 48 pixels are: $r_{top} = (x, y + 23)$, $r_{bottom} = (x, y - 24)$, $r_{left} = (x - 24, y)$ and $r_{right} = (x + 24, y)$. Similarly, to extract ROIs of size 64 × 64 pixels, the offsets used are: $r_{top} = (x, y + 31)$, $r_{bottom} = (x, y - 32)$, $r_{left} = (x - 32, y)$ and $r_{right} = (x + 31, y)$. The same procedure applies for the offsets of other ROI sizes.

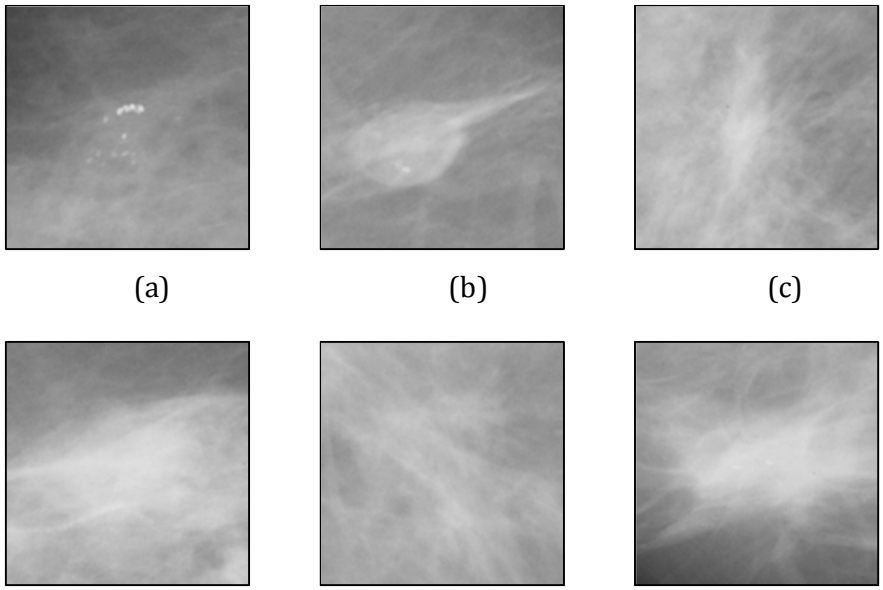


Figure 5.23: ROIs of benign abnormalities (from labeled GT data)
(a) Calcification (b) Circumscribed mass
(c) Spiculated mass (d) Ill-defined mass
(e) Architectural distortion (f) Asymmetrical mass

Malignant and benign ROIs extracted from the acquired mammography images using the technique in Figure 5.22 are shown in Figure 5.23 and Figure 5.24. As observed from Figures 5.23 and 5.24, the ROIs with bright spots (or small clusters) indicate MCCs and the shape of the masses (lesions) in the ROIs is categorized using the BI-RADS lexicon presented in Section 2.6.4.1. The total number of ROIs extracted using the GT data with the acquired mammography datasets is shown in Table 5.2.

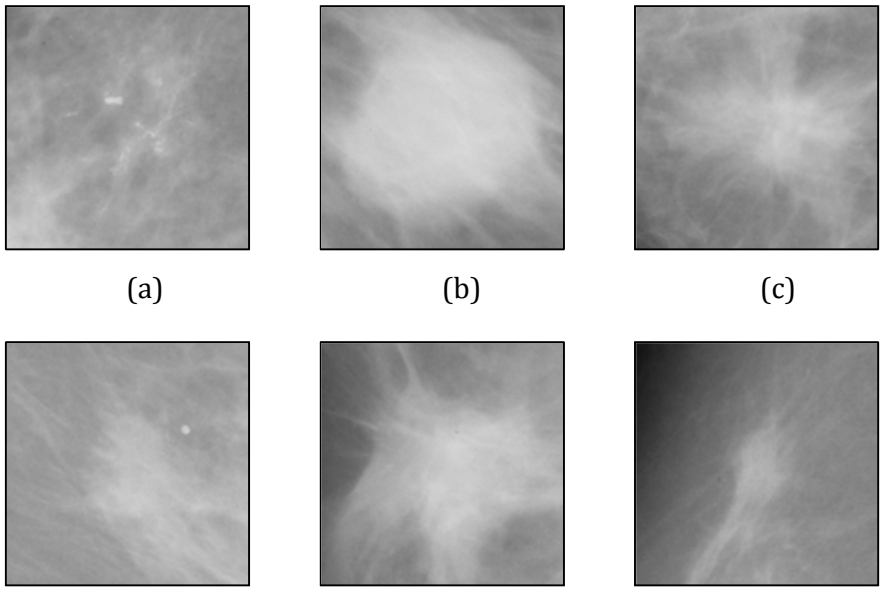


Figure 5.24: ROIs of malignant abnormalities (from labeled GT data)
(a) Calcification (b) Circumscribed mass
(c) Spiculated mass (d) Ill-defined mass
(e) Architectural distortion (f) Asymmetrical mass

5.4.1.1 Necessity for Mammogram Image Processing

For the purpose of the ROI extraction (see Figure 5.22), the necessity for mammogram image processing and segmentation as discussed in Sections 5.3.1 and 5.3.2, arises. One of the main concerns during the ROI extraction is that a

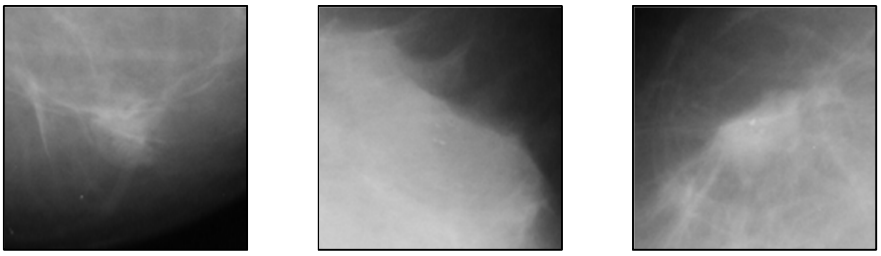
minority of ROIs include the background region (black pixels), which can bias the texture feature extraction process. This happens in mammograms normally when benign and malignant masses/MCCs are present near the edges of the segmented breast profile as indicated in Figure 5.25.

Table 5.2: ROIs extracted from acquired mammography datasets in Table 5.1.

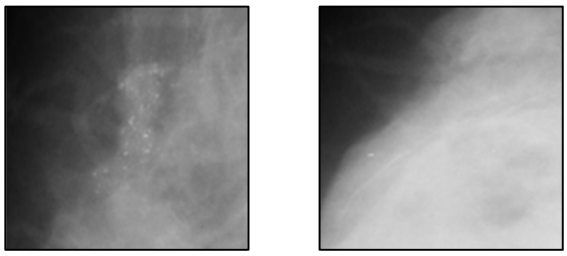
Data Source	Malignant ROIs	Benign ROIs	Database ROIs
University Malaya Medical Centre (UMMC)	64	48	**112**
mini-MIAS Database of Mammograms (Suckling *et al*., 2004)	54	66	**120**
Total ROIs (Samples)	118	114	**232**

In order to extract optimum texture features using GLCMs from the ROIs in Figure 5.25, background pixels in the ROIs need to be separated and the pectoral muscle needs to be segmented from the breast profile region. The theoretical foundation and the practical implementation using algorithms and techniques for background separation and pectoral muscle segmentation have been thoroughly described in Section 3.2 and presented in Sections 5.3.1 and 5.3.2.

In this research, for the purpose of texture feature extraction, only the segmented breast profile region in the ROIs, indicated in Figure 5.26 is used. The ROIs in Figure 5.26 show that the background region (pixels with an intensity value of 0) in the breast profile is excluded from the texture feature extraction process. The reason why the background region is excluded from texture feature extraction is because GLCMs are restricted to perform calculation for counting occurrences of image pixels having an intensity value of 0.



(a) (b) (c)



(d) (e)

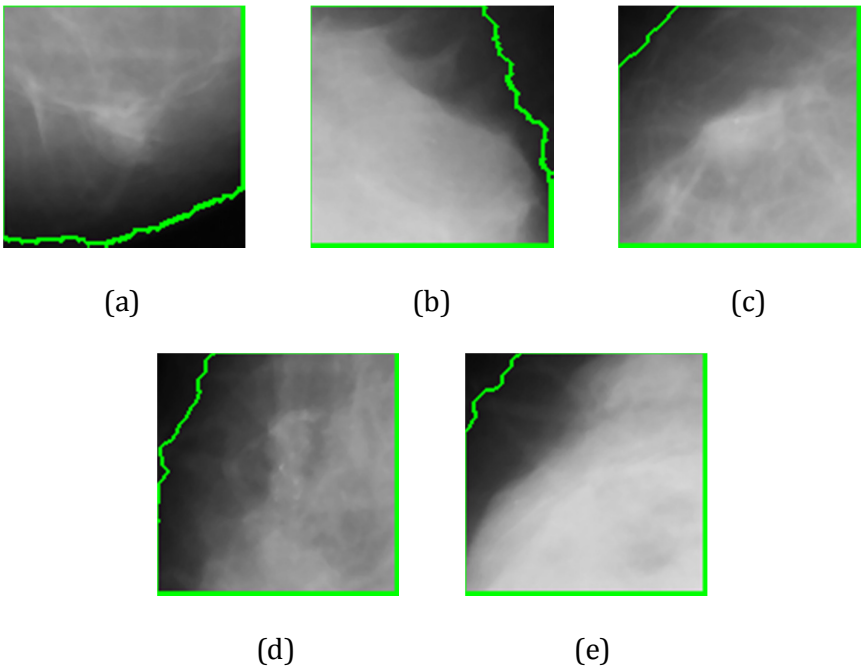
Figure 5.25: ROIs containing unsegmented background region (from labeled GT data)
(a) Benign ROI (b) Benign ROI (c) Malignant ROI
(d) Malignant ROI (e) Malignant ROI

5.4.2 Texture Feature Extraction

For the purpose of texture feature extraction the malignant and benign ROIs obtained in Figure 5.23 and Figure 5.24 are used. GLCMs are known to be the most common and successful techniques for texture analysis of digital mammograms as illustrated in Table 3.3 and Section 3.5.3.

The theoretical background of GLCMs is presented in Section 3.5.3.1.1 of this thesis, whereas Section 3.5.3.1.2 presents the standard GLCM texture descriptors proposed by Haralick *et al.* (1979) as shown in Table 3.4. In this research apart from using standard GLCM texture descriptors discussed by (Haralick, 1973), other recent GLCM texture descriptors discussed by Clausi (2002), Soh &

Tsatsoulis (1999) and the MATLAB Image Processing Toolbox are adopted for texture feature computation, as shown in Tables 5.3 through 5.5 respectively.



(a) (b) (c)

(d) (e)

Figure 5.26: ROIs in Figure 5.25 with segmented background region
(a) Benign ROI (b) Benign ROI
(c) Malignant ROI (d) Malignant ROI (e) Malignant ROI

Table 5.3: GLCM texture descriptors from Clausi (2002)

No.	Texture Descriptor	Formula	Equation No.
1.	Inverse Difference Normalized	$\sum \frac{c(i,j)}{1+\lvert i-j \rvert}$ where $c(i,j)$ is the co-occurrence probability between grey levels (i and j) defined as: $c(i,j) = \frac{p(i,j)}{\sum_{i,j=1}^{G} p(i,j)}$	(5.9)
2.	Inverse Difference Moment Normalized	$\sum \frac{C_{ij}}{1+(i-j)^2}$	(5.10)

Table 5.4: GLCM texture descriptors from Soh & Tsatsoulis (1999)

No.	Texture Descriptor	Formula	Equation No.
1.	Autocorrelation	$\sum_i \sum_j (ij) p(i,j)$ where $p(i,j)$ represents the number of occurrences of grey levels (i and j).	(5.11)
2.	Cluster Prominence	$\sum_i \sum_j \left(i + j - \mu_x - \mu_y\right)^4 p(i,j)$ where $p(i,j)$ is the (i,j)th entry in a normalized GLCM. The mean for the rows and columns of the matrix are: $\mu_x = \sum_i \sum_j i \cdot p(i,j)$ $\mu_y = \sum_i \sum_j j \cdot p(i,j)$	(5.12)
3.	Cluster Shade	$\sum_i \sum_j \left(i + j - \mu_x - \mu_y\right)^3 p(i,j)$	(5.13)
4.	Dissimilarity	$\sum_i \sum_j \lvert i - j \rvert \cdot p(i,j)$	(5.14)
5.	Homogeneity (Soh & Tsatsoulis)	$\sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$	(5.15)
6.	Maximum Probability	$\underset{i,j}{\mathrm{MAX}}\, p(i,j)$	(5.16)

Table 5.5: GLCM texture descriptors from the MATLAB Image Processing Toolbox

No.	Texture Descriptor	Formula	Equation No.
1.	Correlation (MATLAB)	$\sum_{i,j} \frac{(i-\mu_x)(j-\mu_y)p(i,j)}{\sigma_x \sigma_y}$ where $p(i,j)$ is the (i,j)th entry in a normalized GLCM. The standard deviations for the rows and columns of the matrix are: $\sigma_x = \sum_i \sum_j (i - \mu_x)^2 \cdot p(i,j)$ $\sigma_x = \sum_i \sum_j \left(j - \mu_y\right)^2 \cdot p(i,j)$ where μ_x and μ_y are the means for the rows and columns of the matrix respectively	(5.17)
2.	Homogeneity (MATLAB)	$\sum_{i,j} \frac{p(i,j)}{1+\lvert i-j \rvert}$	(5.18)

The GLCM texture feature descriptors shown in Table 5.3 (Clausi, 2002) and Table 5.4 (Soh & Tsatsoulis, 1999) are used for texture feature extraction, since these texture descriptors are more recently proposed in the literature and have indicated promising results in pattern classification problems using textures. The MATLAB Image Processing Toolbox possesses texture analysis capabilities. The *graycopprops* function in MATLAB calculates the statistics specified from GLCMs, which has four texture descriptors, namely: contrast, correlation, energy and homogeneity. It is noticed that the calculation formulae of the *Contrast* and *Energy* texture descriptors in MATLAB Image Processing Toolbox is similar to the ones proposed by Haralick (1973) as listed in Table 3.4. Thus, the remaining two texture descriptors, *correlation* and *homogeneity* are applied in this research as shown in Table 5.5.

In total, 24 GLCM texture descriptors are identified in this research, which are given in Tables 3.4, 5.3, 5.4 and 5.5, corresponding to equations (3.20) to (3.33) and equations (5.9) to (5.18) respectively. The GLCM computational parameters: *Number of grey levels*, *Distance between pixels* and *Angle* (see Section 3.5.3.1.1) used for texture feature extraction are discussed as follows:

(a) *Number of Grey Levels*

All GLCM quantization levels i.e., 8, 16, 32, 64, 128 and 256 are used for the purpose of GLCM texture feature extraction from the grayscale ROIs. This indicates that for each ROI six different grey levels are computed.

(b) *Distance between pixels*

GLCMs are constructed by identifying neighboring pairs of image cells with a distance d from each other and incrementing the matrix position corresponding to the grey level intensity of both cells as indicated in Section 3.5.3.1.1. In this research, the value of d is chosen as $d = 1$, in order to represent the distance between the pixel of interest and the neighboring pixels in each ROI.

(c) *Angle*

In GLCMs it is necessary to define the *direction* of the pair of pixels. The most common GLCM directions for a given distance d are: $P(0^{\circ}, d)$, $P(45^{\circ}, d)$, $P(90^{\circ}, d)$, $P(135^{\circ}, d)$ as indicated in equations (3.15) to (3.18). These four directions, i.e., 0°, 45°, 90°, 135° and their symmetric equivalents, -180°, -135°, 90° and -45° are valid GLCM directions. In this research, all eight GLCM directions, i.e. 0°, 45°, 90°, 135°, -180°, -135°, 90° and -45° are computed for each ROI.

The total number of GLCM texture features computed using the GLCM parameters mentioned above, are reported in Table 5.6. For each ROI, 1152 feature values are computed using the 24 texture descriptors. This means, for each GLCM texture descriptor, 48 feature values are calculated. The GLCM texture features calculated for all the malignant ROIs are shown in Appendix A, Figure A.1, where *rows* represent each ROI sample and *column*s indicate the 1152 texture features for each ROI sample.

Table 5.6: GLCM texture features calculated for each ROI sample

No.	GLCM Parameters	No. of Texture Features
1.	Texture Descriptors (Table 3.4, 5.3, 5.4 and 5.5)	24
2.	GLCM directions (0°, 45°, 90°, 135°, -180°, -135°, 90° and -45°)	8
3.	GLCM distance between pixels ($d = 1$)	1
4.	Number of GLCM grey (quantization) levels: 8, 16, 24, 64, 128, 256	6
	Texture feature values calculated for each ROI:	**1152**

5.4.3 Texture Feature Selection

Feature selection needs to be performed in order to select the optimal subset of features from the 1152 features values obtained in Table 5.6 for each ROI. The optimal subset of selected features will be used to model a SVM classification engine for the purpose of pattern classification. As discussed in Section 4.4 of this thesis, for the purpose of feature selection, a Recursive Feature Elimination (RFE) technique is applied, namely SVM-RFE. The SVM-RFE technique discussed in Section 4.4.2 uses F-scores together with the Random Forest (RF) and the SVM. This feature selection technique is referred to as "*F-score + RF + SVM*" (Chen & Lin, 2006). In this technique, F-scores are used for the purpose of ranking features whereas the RF eliminates unimportant features. This algorithm is implemented in python script, named as *fselect.py* (Feature selection tool for LIBSVM in Python, 2010) by Chen & Lin (2006).

The *fselect.py* tool is evaluated with the 1152 texture features computed from all malignant and benign ROIs (232 samples). Malignant and benign samples represent two different classes of samples. Firstly, the F-score technique presented in Section 4.4.1 is used for ranking all 1152 feature values using equation (4.54). The results obtained for ranking the 1152 features are shown in

Appendix A in Figure A.2. The maximum F-score value computed is 2.6425 whereas the minimum F-score value computed is 0.000415 (considered as 0). The larger the F-score value is, the more the likelihood of that feature being more discriminative.

Table 5.7: GLCM texture descriptors used to select the optimum subset of 1056 features

No.	GLCM Texture Descriptor	Reference	Equation No.	Table No.
1.	**Autocorrelation**	(Soh & Tsatsoulis, 1999)	(5.11)	5.4
2.	**Contrast**	(Haralick, 1973)	(3.21)	3.4
3.	**Correlation (MATLAB)**	-	(5.17)	5.5
4.	**Correlation (Haralick)**	(Haralick, 1973)	(3.22)	3.4
5.	**Cluster Prominence**	(Soh & Tsatsoulis, 1999)	(5.12)	5.4
6.	**Cluster Shade**	(Soh & Tsatsoulis, 1999)	(5.13)	5.4
7.	**Dissimilarity**	(Soh & Tsatsoulis, 1999)	(5.14)	5.4
8.	**Angular Second Moment: Energy**	(Haralick, 1973)	(3.20)	3.4
9.	**Entropy**	(Haralick, 1973)	(3.28)	3.4
10.	**Homogeneity (MATLAB)**	-	(5.18)	5.5
11.	**Homogeneity (Soh & Tsatsoulis)**	(Soh & Tsatsoulis, 1999)	(5.15)	5.4
12.	**Maximum Probability**	(Soh & Tsatsoulis, 1999)	(5.16)	5.4
13.	**Sum of Squares: Variance**	(Haralick, 1973)	(3.23)	3.4
14.	**Sum Average**	(Haralick, 1973)	(3.25)	3.4
15.	**Sum Variance**	(Haralick, 1973)	(3.26)	3.4
16.	**Sum Entropy**	(Haralick, 1973)	(3.27)	3.4
17.	**Difference Variance**	(Haralick, 1973)	(3.29)	3.4
18.	**Difference Entropy**	(Haralick, 1973)	(3.30)	3.4
19.	**Information Measure of Coefficient 1**	(Haralick, 1973)	(3.31)	3.4
20.	**Information Measure of Coefficient 2**	(Haralick, 1973)	(3.32)	3.4
21.	**Inverse Difference Normalized**	(Clausi, 2002)	(5.9)	5.3
22.	**Inverse Difference Moment Normalized**	(Clausi, 2002)	(5.10)	5.3

Using these F-scores of the 1152 features, the technique presented in Step (2) of Section 4.4.2 (Chen & Lin, 2006), performs 10-fold CV using the RF technique to filter out unimportant features. The result of this process is an optimal subset of features. Applying this technique, an optimal subset of 1056 texture features is obtained, as shown in Appendix A in Figure A.3. As the optimum subset of features consists of 1056 texture feature values, this indicates the feature selection process recursively eliminates 96 feature values. The 96 feature values filtered correspond to two GLCM texture descriptors, namely the *Inverse Difference Moment* (in equation (3.24)) and the *Maximum Correlation Coefficient* (in equation (3.33)). Thus, the optimum subset of 1056 texture features shown in Appendix A in Figure A.3 is obtained using the 22 GLCM texture descriptors shown in Table 5.7.

5.4.4 Feature Normalization

Prior to development of the SVM classification engine, the optimal subset of 1056 texture features obtained in Section 5.4.3, need to be represented in a normalized scale. In order for the feature data to fit the SVM properly, all 1056 features are scaled (normalized) in the range between 0 and 1. Feature normalization is performed using the following expression:

$$NF(x) = \frac{F(x) - \min(F(x))}{\max(F(x)) - \min(F(x))} \tag{5.19}$$

where $F(x)$ for $x = 1,2,3, \ldots, 1056$ represents the feature of interest and $\min(F(x))$ and $\max(F(x))$ represent the minimum and maximum values corresponding to the feature of interest $F(x)$. The feature data in Appendix A in Figure A.4 is normalized in the range between 0 and 1 as shown in Figure A.5.

5.5 Classification Engine Development

5.5.1 Feature Labeling and Adjustment

In order for the feature data for SVM training and testing the data needs to be adjusted in and presented in a proper format to LIBSVM (Chang & Lin, 2010). In order to serve this purpose, all normalized features values are labeled, where labels are represented using integers. Normalized feature values corresponding to their labels are mathematically represented by the matrix R, in the form:

$$R = \begin{bmatrix} l_{11}:x_{11} & \cdots & l_{1k}:x_{1k} & \cdots & l_{1c}:x_{1c} \\ l_{21}:x_{21} & \cdots & l_{2k}:x_{2k} & \cdots & l_{2c}:x_{2c} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{p1}:x_{p1} & \cdots & l_{pk}:x_{pk} & \cdots & l_{pc}:x_{pc} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{r1}:x_{r1} & \cdots & l_{rk}:x_{rk} & \cdots & l_{rc}:x_{rc} \end{bmatrix} \tag{5.20}$$

where l is an integer representing the feature label for $l = \{1,2,\dots,1056\}$, x represents the normalized feature value, c indicates the number of texture features for $c = \{1,2,\dots,1056\}$ and r indicates the number of ROI samples for $r = \{1,2,\dots,232\}$.

During feature labeling the samples belonging to the malignant and benign classes are labeled using the information from the GT data (in Section 5.2.1). Malignant samples are represented as the $+ve\ class$ or $class{-}1$ and benign samples are represented as the $-ve\ class$ or $class{-}2$. The LIBSVM feature file obtained after associating the class memberships of all the samples is shown as Appendix A in Figure A.6., where the first column represents the class label for each sample (row).

5.5.2 Training and Testing Data Separation

To implement the Non-linear SVM (in Section 4.2.4), the normalized feature needs to be separated into two distinct sets, i.e. the training set and the testing/validation set. As observed from Table 5.2, the total number of ROI samples obtained from the acquired mammography data is 232, out of which 118 are malignant samples ($class{-}1$) and the remaining 114 are benign samples ($class{-}2$).

In order to split the feature data into the training and testing sets, the Holdout method presented in Section 4.1.3.1 is adopted, where two-third (70 percent) of the samples from both classes are allocated to the training set and the remaining one-third (30 percent) of the samples from both classes are allocated to the testing set. The specification of the samples in training and testing sets is indicated in Table 5.8.

Table 5.8: Ratio of samples used for training and testing from the UMMC and MIAS datasets

Class	Number of Samples	Training Set (70% Samples)	Testing Set (30% Samples)
Malignant (+ve) ($class{-}1$)	118	82	36
Benign (-ve) ($class{-}2$)	114	80	34
Total Samples:	**232**	**162**	**70**

5.5.3 SVM Model Development

The SVM is implemented in this research using LIBSVM library (Chang & Lin, 2010) integrated into MATLAB. As observed from Table 5.8, a total of 162 samples from both classes are used for SVM training (memorization/learning), while the remaining 70 samples are used for testing the accuracy of the trained model or SVM classification engine for unseen data samples. Since the training accuracy of

the SVM classifier is evaluated using the 10-fold CV approach discussed in Section 4.1.3.2 of this thesis, thus, on each CV fold, training and testing samples are selected randomly, so as to ensure that the developed classification engine does not overfit the training data. The SVM training engine proposed for constructing the classification engine and performing SVM hyperparameter optimization is illustrated in Figure 5.27.

5.5.3.1 SVM Parameter Optimization

The training or learning accuracy of the SVM classification engine is estimated by tuning the error penalty parameter, C (in equation (4.29)). In this research, the RBF (*Gaussian*) kernel (in equation (4.44)) is used with Non-linear SVM as discussed earlier in Section 4.2.4. The parameter γ in the RBF kernel which controls the width of the Gaussian needs to be optimized with respect to the SVM hyperparameter C. Thus, two SVM hyperparameters (C, γ) need to be determined in order to construct a classifier with an optimum balance between its memorization and generalization capability.

The *Grid Search* method proposed by Hsu *et al.* (2003) and discussed in Section 4.2.4 is used in this research for SVM hyperparameter optimization, as indicated in Figure 5.27. In the Grid Search method, exponentially growing sequences of parameters (C, γ) are used to identify optimum parameter values with respect to the best 10-fold CV accuracy. In this trial and error procedure, sequences of parameters in the range, $C = [2^1, 2^2 \ldots, 2^{20}]$ and $\gamma = [2^{-20}, 2^{-19}, \ldots, 2^1]$ are evaluated for 100 × 100 = 10,000 combinations respectively.

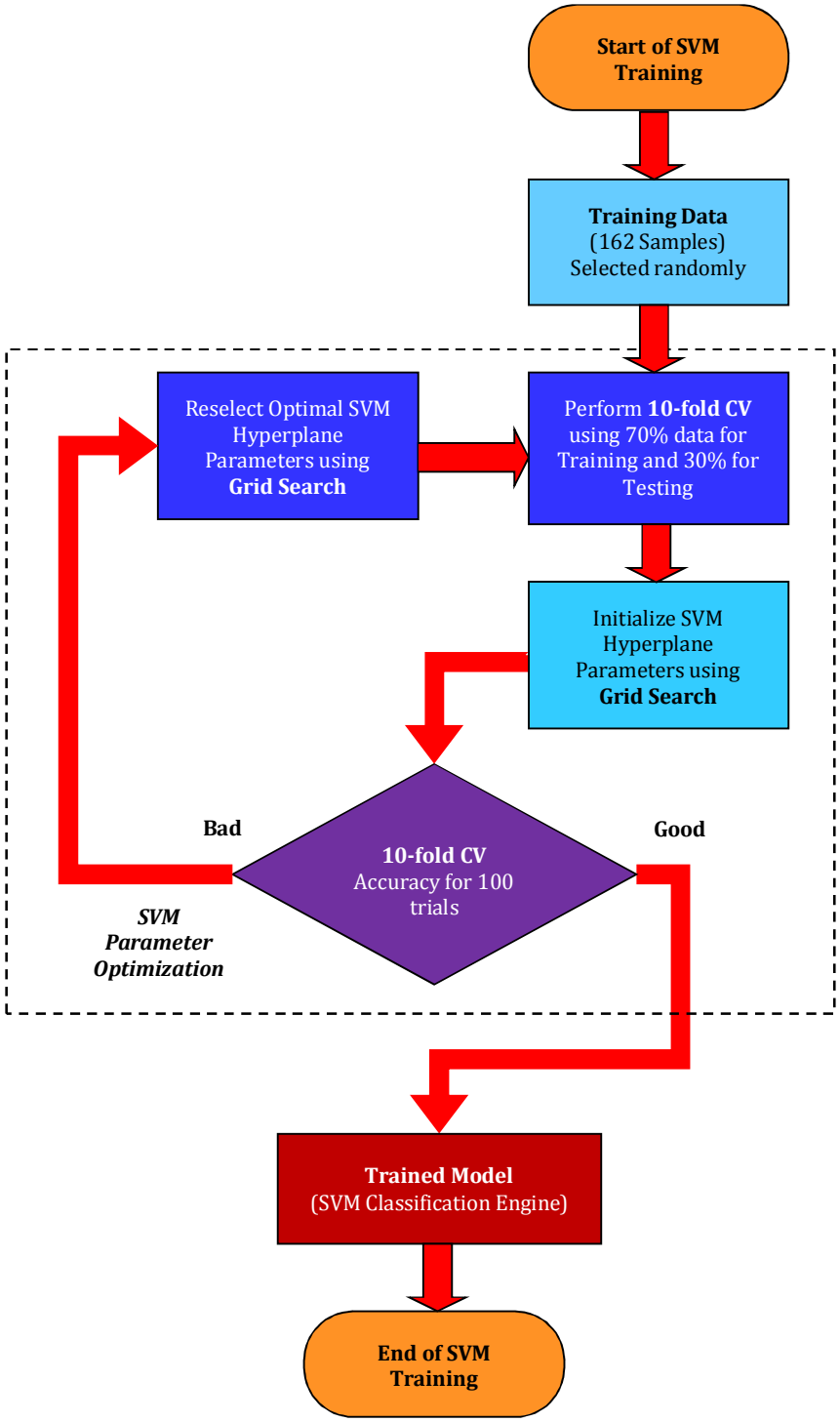


Figure 5.27: The SVM training engine proposed for constructing the classification engine and performing hyperparameter optimization

For each pair of (C, γ) the 10-fold CV performance of the trained model is measured by splitting the training data (162 samples) into two smaller CV subsets, using the Holdout method presented in Section 4.1.3.1. Using this method, the CV training subset contains two-thirds of the training samples (113 samples), whereas the CV testing subset contains the remaining one-third of the training samples (49 samples). The selection of the CV training and CV testing subsets is repeated 100 times for 10-fold CV trials, where on each trial, the samples used for CV training and CV testing are selected randomly.

Iterating different parameter combinations of (C, γ) experimentally, the optimum SVM parameters using Grid Search are found to be: $C = 64$ and $\gamma = 0.001953125$, which obtain the highest 10-fold CV accuracy of 87.83 percent as indicated in Figure 5.28. The training (memorization) accuracy[1] of the SVM classification engine is calculated using the following expression:

$$Training\ Accuracy = \frac{T_C}{T_S} \times 100\% \tag{5.21}$$

where T_C represents the total number of samples correctly classified by the SVM and T_S represents the total number of samples used for CV testing. As 49 samples are used for CV testing ($T_s = 49$), the 10-fold CV results obtained indicate that 48 out of 49 samples are classified correctly by the SVM, thus, $T_C = 48$. Using equation (5.21) with parameter values $T_C = 48$ and $T_s = 49$ a training accuracy of 97.6 percent is achieved for the SVM classification engine. This indicates that the developed classification engine has good learning capability.

[1]Training accuracy is the measure of the memorization and learning capability of the classifier. Training accuracy is calculated in percentage of the total number of samples used for SVM testing divided by the total number of samples correctly classified by the SVM.

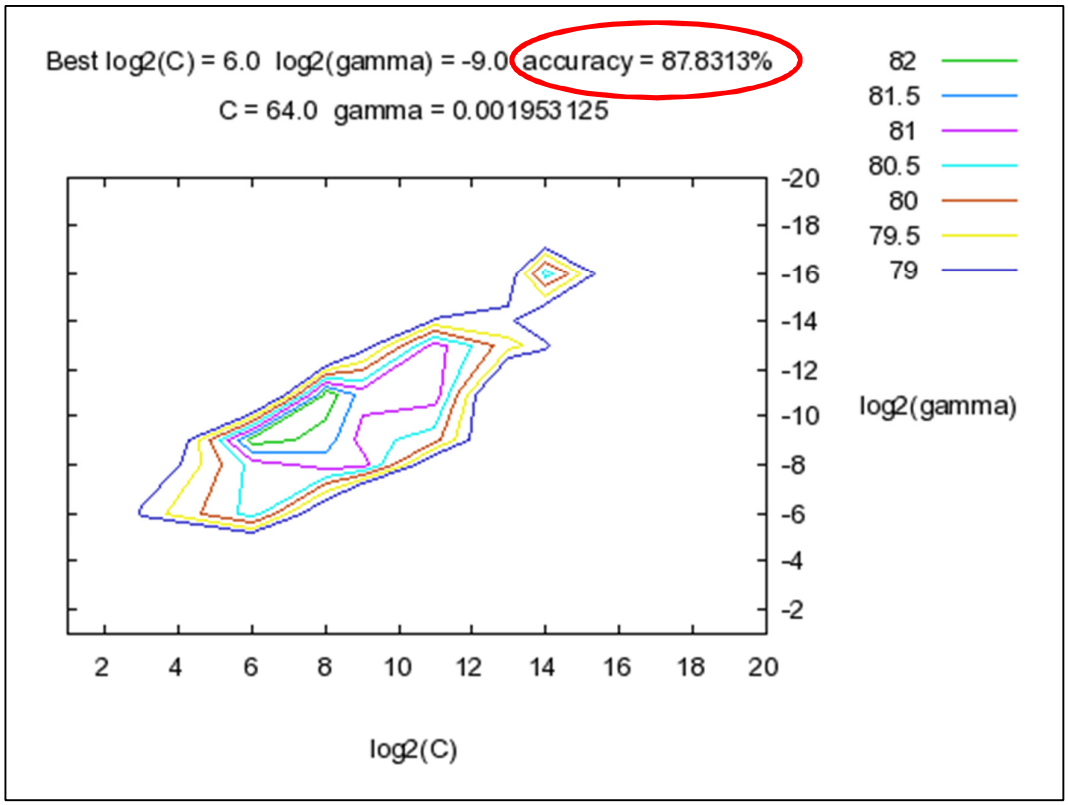


Figure 5.28: Grid Search for selection of optimal of SVM hyperparameters (C, γ)

5.5.3.2 Probability Estimation

Besides conducting classifications, SVMs also compute the probabilities for each class (Wu *et al.*, 2004). This supports the analytic concept of generalization and certainty. Given that r_{ij} is an estimate for the probability of the output of pairwise classifiers between class i and class j (i.e., $r_{ij} \approx \mathrm{P}(y = \{i, j\}, x), r_{ij} + r_{ji} = 1$) and that p_i is the probability of the i^{th} class, the probability $p = (p_1, \dots, p_k)$ of a class can be derived via a quadratic programming (QP) problem (Oskoei & Hu, 2008):

$$\min_p \sum_{i=1}^{k} \sum_{j:j\neq 1} \left(r_{ij}\, p_j - r_{ji}\, p_i\right)^2, \ \sum_{i=1}^{k} p_i = 1 \qquad p_i \geq 0, \ \forall i \tag{5.22}$$

The pairwise probability information defined in equation (5.22) is computed using the LIBSVM library (Chang & Lin, 2010), to estimate the probabilities of the tested

samples. The probability estimates (decision values) for the testing set, provides additional information for selection of samples with higher confidence measures (probabilities), as will be discussed in Section 5.5.3.5 later.

5.5.3.3 SVM Training

SVM training is performed by integrating the LIBSVM library (Chang & Lin, 2010) into MATLAB. After obtaining the optimal pair of SVM hyperparameters (C, γ) (see Section 5.5.3.1), the SVM is trained using the 162 training samples as indicated in Table 5.8 and Figure 5.27. The LIBSVM MATLAB executable, *svmtrain.mexw32* is employed for SVM training as shown in Appendix B in Figure B.1.

The optimized SVM hyperparameters (C, γ) obtained from the Grid Search method in Section 5.5.3.1 are used to model a Non-linear SVM for binary classification, as shown in Appendix B in Figure B.2. In Figure B.2, *samples.txt* represents the 162 training samples in Table 5.8. The '-v 10' parameter used in the training string performs 10-fold CV, where the result obtained is the memorization accuracy of the trained classifier. During the SVM training, the SMO process finishes on the 5494^{th} iteration obtaining the highest 10-fold CV accuracy of 87.83 percent as indicated in Section 5.5.3.1 and Figure 5.28.

The trained model generated in MATLAB after SVM training is shown in Appendix B in Figure B.3, where the SVM *model* parameters are shown in Figure 5.29. As observed from Figure 5.29, the trained model contains 52 support vectors (SVs), defined by the constraint $0 \leq \alpha_i \leq C$ in equation (4.34), where the malignant class $(class{-}1)$ has 29 SVs and the benign class $(class{-}2)$ has 23 SVs. The total number of bounded SVs (BSVs) computed by the model is 10, justifying that the condition

$\alpha_i = C$ (in equation (4.35)) is true. The rho parameter defined as $rho = -b$ in the decision function in equation (4.42) is computed to be: $rho = 13.3317$.

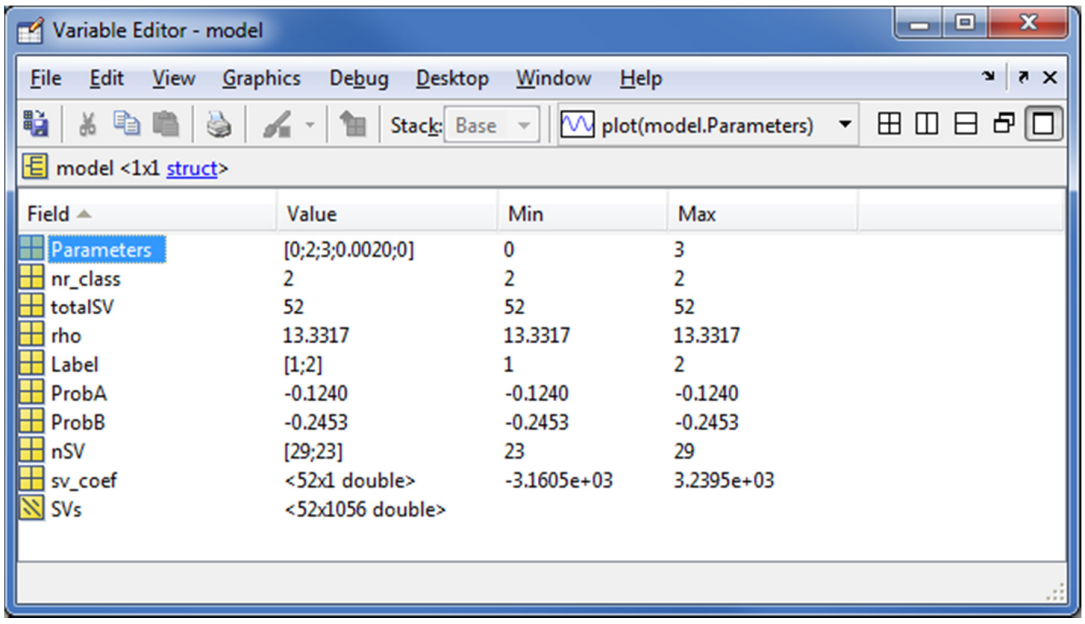


Figure 5.29: SVM classification engine – Trained model

The separating boundaries between the two classes (malignant and benign) of the training data in the SVM classification engine (in Figure 5.29) are shown in Figure 5.30. This plot is drawn using the "SVM Toy" tool, *svm-toy.exe* included in the LIBSVM (Chang & Lin, 2010) library. As observed from Figure 5.30, the purple (dark) dots represent malignant samples, while the green (light) dots represent benign training samples. A training sample in the purple region indicates that sample is classified as malignant ($class—1$), whereas a sample in the green region indicates that sample is classified as benign ($class—1$). Observing the non-linear soft-margin boundaries between the two classes of training data, it is suitable to imply that samples in both classes are well separated (with a only few misclassifications), which indicates that the trained model has good learning and memorization capability,

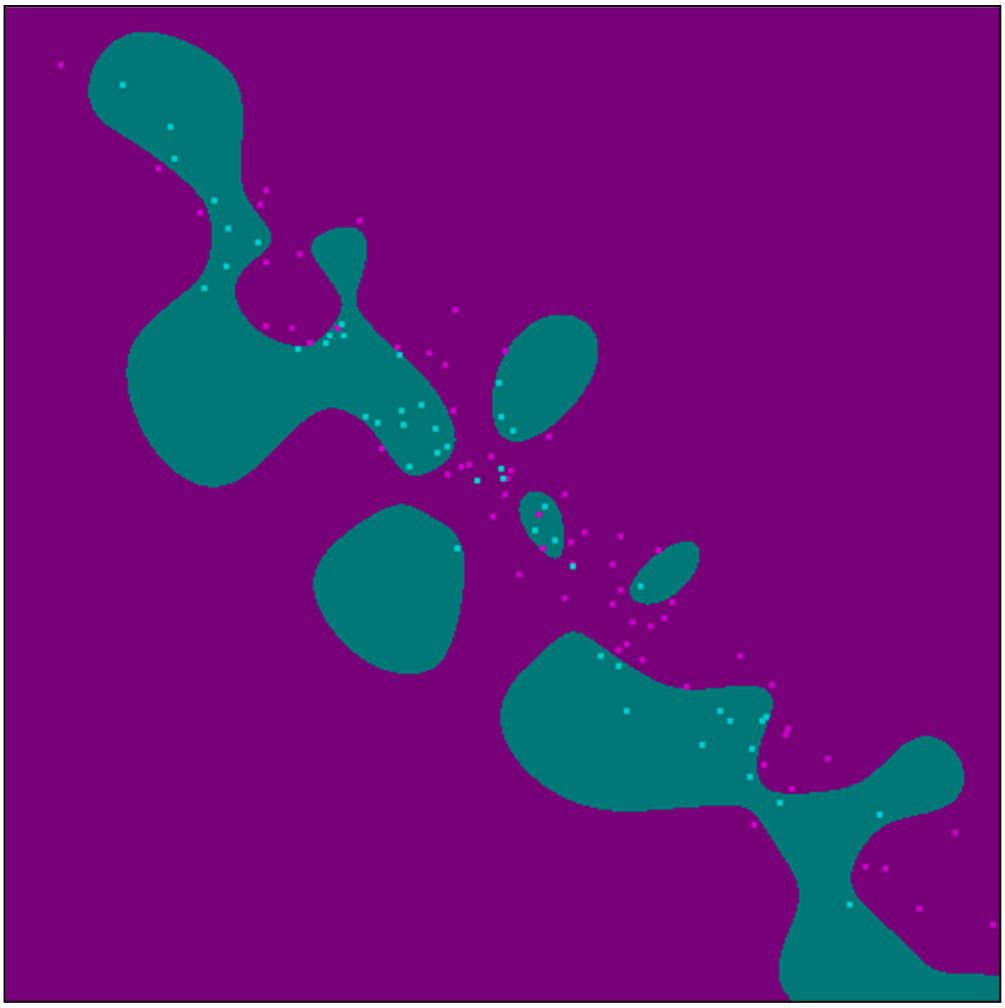


Figure 5.30: Separating boundaries of the SVM classification engine in Figure 5.29

5.5.3.4 SVM Testing and Validation

During the testing and validation phase, the classification engine developed in Section 5.5.3.3 is used with the remaining 30 percent testing samples (70 samples) as shown in Table 5.8 to test the accuracy of the developed system.

The LIBSVM MATLAB executable for SVM testing, *svmpredict.mexw32* is evaluated for testing and validating the classification engine, as shown in Appendix B in Figure B.4. The samples used for SVM testing are in the exact same format as the training data samples shown in Appendix A in Figure A.6. The class membership

associated with the SVM testing samples can be taken as either class, malignant ($class{-}1$) and/or benign ($class{-}2$). This is because the class labels are only useful for computing the k–fold CV accuracy during SVM training, and since the labels of the testing samples in are unknown, they can be considered as either class. In this research all SVM testing samples are represented as benign ($class{-}2$). So, during testing the class labels are ignored as they going to be predicted by the SVM classification engine.

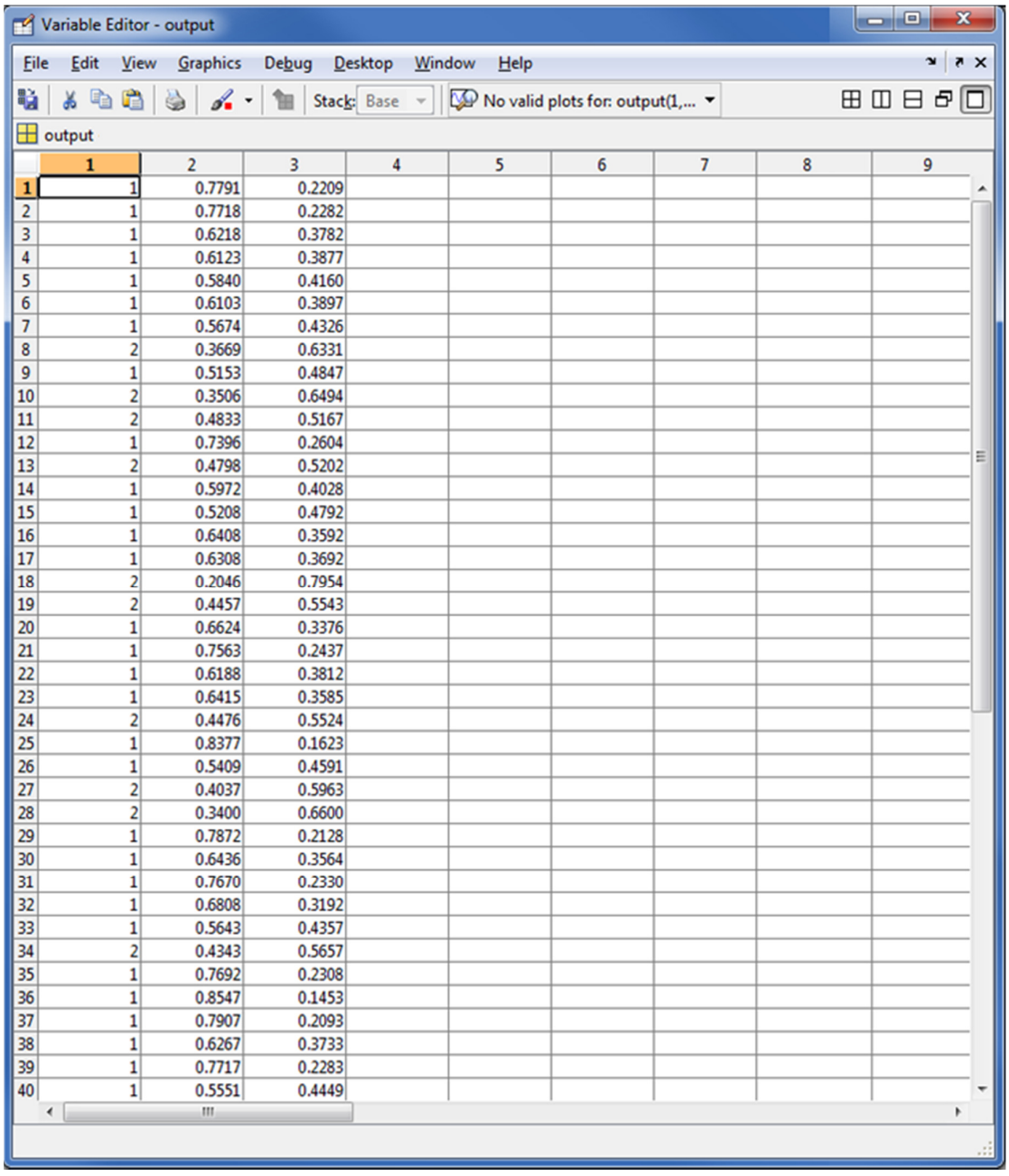
SVM testing and validation is performed in MATLAB using the SVM classification engine and the testing samples as shown in Appendix B in Figure B.5. The '-b 1' parameter used in the SVM prediction string computes the probability estimates of the tested samples, as discussed in Section 5.5.3.5. The classification accuracy of the SVM for testing the 70 samples (in Table 5.8) is found to be 97.14 percent, which indicates that out of 70 samples 68 samples have been correctly classified by the SVM classification engine.

The resulting parameters '*outputlabel*' and '*probability*' in Figure B.5 contain the classification results as the 'predicted class labels' and 'probability estimates' respectively for the tested samples, as shown in Figure 5.31. As observed from Figure 5.31, the first column represents the predicted class labels, i.e. $class{-}1$ or $class{-}2$ for the tested samples, whereas the second and third columns represent the probability of $class{-}1$ (p_{class1}) and $class{-}2$ (p_{class2}) for each testing sample.

5.5.3.5 Logic System for False Positive (FP) Reduction

The SVM testing and validation results as shown in Figure 5.31 include probability estimates of the tested samples. These probability estimates are used to model a

decision-logic system, to reduce the number of FPs; serving as the one of the key objectives outlined in this research in Sections 1.2 and 1.3.



	1	2	3	4	5	6	7	8	9
1	1	0.7791	0.2209						
2	1	0.7718	0.2282						
3	1	0.6218	0.3782						
4	1	0.6123	0.3877						
5	1	0.5840	0.4160						
6	1	0.6103	0.3897						
7	1	0.5674	0.4326						
8	2	0.3669	0.6331						
9	1	0.5153	0.4847						
10	2	0.3506	0.6494						
11	2	0.4833	0.5167						
12	1	0.7396	0.2604						
13	2	0.4798	0.5202						
14	1	0.5972	0.4028						
15	1	0.5208	0.4792						
16	1	0.6408	0.3592						
17	1	0.6308	0.3692						
18	2	0.2046	0.7954						
19	2	0.4457	0.5543						
20	1	0.6624	0.3376						
21	1	0.7563	0.2437						
22	1	0.6188	0.3812						
23	1	0.6415	0.3585						
24	2	0.4476	0.5524						
25	1	0.8377	0.1623						
26	1	0.5409	0.4591						
27	2	0.4037	0.5963						
28	2	0.3400	0.6600						
29	1	0.7872	0.2128						
30	1	0.6436	0.3564						
31	1	0.7670	0.2330						
32	1	0.6808	0.3192						
33	1	0.5643	0.4357						
34	2	0.4343	0.5657						
35	1	0.7692	0.2308						
36	1	0.8547	0.1453						
37	1	0.7907	0.2093						
38	1	0.6267	0.3733						
39	1	0.7717	0.2283						
40	1	0.5551	0.4449						

Figure 5.31: SVM testing and classification results using LIBSVM in MATLAB

As shown in Figure 5.31, each row represents a testing sample, where the first column is the predicted label of the tested sample. The second and third columns

represent the probability estimates of the malignant samples ($+ve\ class$) and benign samples ($-ve\ class$) respectively. The probability estimates indicate the belongingness of a testing sample to a class. For a testing sample, if the probability of malignant is higher than benign, then that sample is classified as malignant while if the probability of benign is higher than malignant then that sample is classified as benign. This indicates that the class with the higher probability will be predicted class label. The probability estimates for each sample are computed such that:

$$p_{class1} + p_{class2} = 1 \tag{5.23}$$

As observed from equation (5.23), summing the malignancy and benign probability values p_{class1} and p_{class2} for any testing sample, equals to 1. In order to model a decision-logic system to reduce the number of FPs, the differences between the probabilities for both classes is calculated for each testing sample, using the following expression:

$$D = |p_{class1} - p_{class2}| \tag{5.24}$$

The probability differences (computer using equation (5.24)) of the misclassified samples are compared with the probability differences of the correctly classified samples in order to determine a threshold, which is used to reduce the FPF. Inspection of the probability difference data from the 70 testing samples shows that a threshold value of $t = 0.08$ can reduce the FPs. A pseudo code of the decision-logic system using equation (5.24) in MATLAB is given in Figure 5.32. The experimental results obtained from testing the decision logic-system are presented and discussed in Section 6.1.4.1.3.

```
%Repeat process for each testing sample
for each testing_sample i
    %Calculate probability difference (threshold)
    t = abs(pclass1(i)-pclass2(i));

    %Reduce FP rate
    if (label(i) == 1) AND (t < 0.08)
        newlabel(i) = 2;
    %If condition not met
    else
        newlabel(i) = label(i);
    end
end
```

Figure 5.32: Decision-logic system for reduction of false positives (FPs)

5.6 Summary

This chapter presented the modeling of the framework (system) proposed in Chapter 1 for the classification of benign and malignant abnormalities in digital mammograms. As discussed in Chapter 3, the proposed framework is composed of two main techniques, namely, image processing and machine learning. The image process and machine learning techniques identified in Section 3.3, are applied in the proposed system in this chapter, which are discussed in Section 5.1 and 5.2. The modeling of the system consists of three main stages, namely: Mammogram Image Processing (Section 5.3), Texture Feature Extraction and Selection (Section 5.4) and Classification Engine (Section 5.5). Sections 5.3 through 5.5 describe each stage in detail during the development of the proposed framework.

The preliminary testing results of the developed framework in Section 5.5.3.4 and Section 5.5.3.5 show promising results for the classification of malignant and benign abnormalities in digital mammograms. The framework developed in this chapter is tested thoroughly, where the experimental results are presented and discussed in Chapter 6.

CHAPTER 6

EXPERIMENTAL RESULTS AND DISCUSSION

6.0 Overview

This chapter presents the experimental results of the developed system in Chapter 5. Section 6.1 presents and discusses the SVM training results relative to the memorization and learning of the binary SVM classifier. Section 6.1 also presents and discusses the SVM testing and validation results for unseen samples. In order to perform a comparative research, Section 6.2 presents the experimental results obtained after evaluating the developed framework using different machine learning algorithms other than the SVM. The experimental results of the compared machine learning models are discussed in the last part of Chapter 6.

6.1 Experimental Results of Proposed Framework

6.1.1 Image Segmentation Performance Indices

The accuracy of the mammogram segmentation (Stage 1 in Figure 5.2) algorithm in this research is evaluated by deriving quantitative measures by comparing each segmented mammogram mask with its corresponding *gold standard*. In this research, the gold standard is obtained by manually segmenting the breast region from the background region for all mammogram images acquired. To serve this purpose, the boundary of the breast is traced to extract the real breast region, which results in a *Ground Truth* (GT) image as shown in Figure 6.1.

Quantitative measures using the Receiver Operating Characteristics (ROCs) (see Section 3.3.6) are used to describe the accuracy of the mammogram segmentation process. The region (mask) obtained from the segmentation result which matches the GT image, is denoted as the *True Positive* (TP) pixels, which expresses that the segmentation algorithm has found a portion of the breast. The pixels shown in the GT image but not shown in the mask are denoted as *False Negative* (FN) pixels, which are the missing pixels in the breast region. Finally, the pixels not in the GT image, but in the shown in the mask, are denoted as *False Positive* (FP) pixels.

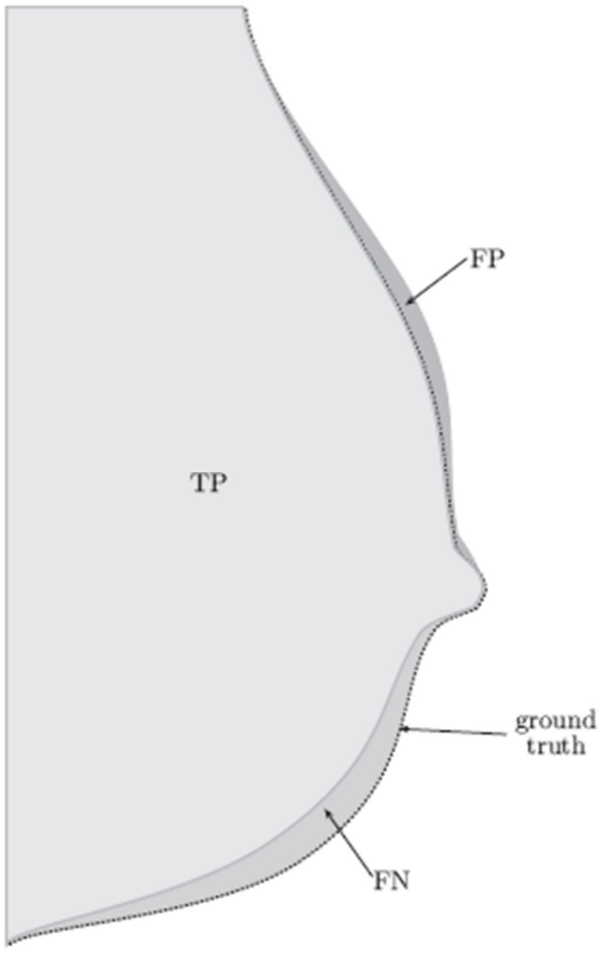


Figure 6.1: Image segmentation performance indices: TP, FP and FN

Using the mammogram segmentation performance indices (TP, FP and FN) in Figure 6.1, two metrics relating to the segmentation performance are derived, namely: *Completeness* (CM) and *Correctness* (CR). In mammogram segmentation, CM is the percentage of the GT region, which describes the segmented region, using the following expression:

$$Completeness\ (\mathrm{CM}) = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FN})} \tag{6.1}$$

CM ranges from 0 to 1, with 0 indicating that none of the regions are properly partitioned, and 1 indicating that all the regions were segmented. For example, a value of CM = 0.92 indicates a 92 percent overlap with the GT image. Similarly, CR represents the percentage of correctly segmented breast region (profile), using the following expression:

$$Correctness\ (\mathrm{CR}) = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FP})} \tag{6.2}$$

Similar to CM, the optimum value for CR is 1 and the minimum value is 0. Lower values of CM indicate *over-segmentation*, whereby a region in the GT is represented by two or more regions in the examined segmented image. Similarly, *under-segmentation* is defined for CR, where two or more regions in the GT are represented by a single region in the segmented image. In mammogram segmentation, the segmentation algorithm is considered accurate if the percentage of CM and CR is greater than 95 percent. A more general measure of the mammogram segmentation performance is achieved by combining CM and CR into a single measure known as *Quality* (Q), using the following expression:

$$Quality\ (\mathrm{Q}) = \frac{\mathrm{TP}}{(\mathrm{FN}+\mathrm{FP}+\mathrm{TP})} \tag{6.3}$$

Similarly, the optimum value for Q is 1 and the minimum value is 0. Results obtained from mammogram preprocessing (in Figures 5.1 and 5.2) indicate some influence on the effectiveness of the segmentation algorithm, but since the noise removal and background/artifact suppression algorithms are not image enhancement algorithms, there are no GT images present. Thus, it is considered non-trivial to quantitatively measure the effective of the mammogram preprocessing.

6.1.1.1 Image Segmentation Results

The mammogram segmentation algorithm (in Figures 5.1 and 5.2) is evaluated on all the 582 mammogram samples as shown in Table 5.1. To demonstrate the robustness of the segmentation algorithm, it has been evaluated on mammograms with differing breast densities such as fatty, fatty fibroglandular and dense fibroglandular tissues.

Segmenting all the 582 mammogram images (in Table 5.1), the average CM and CR obtained are 0.996 and 0.981 respectively, signifying that the mammogram segmentation algorithm is robust with respect to different tissue densities. This implies that the average proportion of the segmented breast region detected by the algorithm is 99.6 percent, while 1.9 percent of the background is mislabeled as the breast region.

With few exceptions the mammogram segmentation algorithm performed well with sufficient reliability to retain the nipple in the breast region (profile). After

segmentation is completed, it is computed that the average breast region contains approximately 208,400 pixels. Thus, on average, each segmented mask misses 366 pixels from the breast region and mistakes 1742 pixels from the background region as breast pixels, which gives a quality of Q = 0.98. The adaptability of the segmentation algorithm in terms of tissue density is illustrated in the three following experiments.

Experiment 1 — Fatty Tissue

The first experiment deals with mammograms which predominantly comprise of fatty tissue. The segmented image closely approximate to the breast region as represented by the GT image. The quantitative measures indicate that the segmented breast regions are marginally under-segmented (2 percent), but do contain the breast region in their entirety. The mean CM and CR values for all fatty tissues are computed to be 0.99 and 0.96 respectively.

Experiment 2 — Fatty-Fibroglandular Tissue

The second experiment deals with mammograms which comprise of fatty fibroglandular tissue. The mean CM and CR values for all fatty fibroglandular tissues are computed to be 1.00 and 0.99 respectively. Since, the CM and CR values are closest to the optimum values that can be obtained, this indicates that the segmentation error is very small, i.e., less than 1 percent.

Experiment 3 — Dense-Fibroglandular Tissue

The final experiment deals with mammograms comprising of dense fibroglandular tissue. The mean CM and CR values for all dense fibroglandular tissues are computed to be 1.00 and 0.98 respectively. For dense fibroglandular tissues the

nipple in the breast profile in all the mammograms is retained and the segmented breast region compares well with the GT images.

As a conclusion to the three experiments above, the mammogram segmentation algorithm is invariant to changes in tissue density. However, no segmentation algorithm can be considered 100 percent robust, this is due to the heterogeneous nature of mammograms. Typical mammogram acquisition problems include: scanner induced artifacts, excessive background noise and scratches and dust which influence the reliability of the segmentation algorithm. The mammogram segmentation results indicate that, for all 582 mammograms evaluated, only 9 mammograms (1.5 percent) fell marginally short of the 95 percent accuracy indicator and 3 mammograms (0.5 percent) were over-segmented, which attributed primarily to indistinct boundaries.

The reason for the 2 percent segmentation inaccuracy is because these 12 mammograms are considered as special cases, which have a highly non-uniform background and less contrast in the area above the breast tissue region. So, the segmentation results in a roughly extracted breast contour corresponding to the breast region. The CM and CR measures of the 9 (1.5 percent) oversegmented mammogram images are computed to be 0.87 and 0.99 respectively, which indicates that the GT image contains the entire segmented region.

6.1.2 Feature Selection Results

Texture features are computed using the GLCMs of all ROI samples (malignant and benign) for the purpose of binary classification using SVM. The feature selection algorithm evaluated in this research is known as "*F-score + RF + SVM*" technique

(Chen & Lin, 2006), which is discussed in detail and the experiments are reported in Section 4.4 and Section 5.4.3 of this thesis respectively.

Initially, 24 GLCM texture descriptors are used for the purpose of feature selection, as indicated in Table 5.6. After feature selection (Chen & Lin, 2006), the optimum subset of texture features is computed to be 1056, which corresponds to 22 GLCM texture descriptors as indicated in Table 5.7. This indicates that the Recursive Feature Elimination (RFE) technique eliminates 2 GLCM texture descriptors corresponding to 96 texture feature values. The optimum subset of 1056 texture features obtains the highest 10-fold CV accuracy of 82.30 percent. The following section discusses the feature selection results obtained using the proposed technique.

6.1.2.1 Discussion of F-score Results

The F-score feature ranking algorithm (in Section 4.4.1) uses a RFE technique, namely the SVM-RFE as discussed in Section 4.4.2. In order to compute the F-scores for the GLCM texture features using the SVM-RFE based approach, binary confusion matrix performance indices (TP, FP and FN) need to be computed at first, as indicated in Table 3.1. Prior to computation of F-scores, the $Precision$ and $Recall$ for each texture feature is computed using the following expressions:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{6.4}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{6.5}$$

Using the precision and recall values obtained from equations (6.4) and (6.5), F-scores for each texture feature are calculated using the following expression:

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6.6}$$

The F-scores computed for the GLCM texture features are shown in Appendix A in Figure A.2, whereby the optimum subset of features selected using the proposed technique is shown in Appendix A in Figure A.3. Since the proposed feature selection algorithm in this research obtains a 10-fold CV accuracy of 82.30 percent using the "F-score + RF + SVM" technique, this indicates that the optimum subset of features selected has a negligible correlation between each feature. This is because during SVM-RFE, features with lower F-scores are eliminated.

6.1.3 SVM Training Validation

After obtaining the optimal pair of SVM hyperparameters (C, γ) (in Section 5.5.4.1), the SVM is trained for binary classification using the 162 training samples as indicated in Table 5.8.

Using the training approach discussed in Section 5.5.3.3, the highest 10-fold CV accuracy of the SVM classification engine obtained is 87.83 percent, as indicated in Figure 5.28. The training accuracy of the classification engine is calculated using equation (5.21), which results in a training accuracy of 97.60 percent. The training accuracy indicates that the developed classification engine has good learning and memorization capability. The separating boundaries (soft-margin) between the two classes of the training data, $class{-}1$ (malignant) and $class{-}2$ (benign) is illustrated in Figure 5.30.

Prior to SVM training, optimum SVM hyperplane parameters (C, γ) need to be determined. As mentioned throughout this thesis, for the purpose of SVM hyperparameter optimization, 10-fold CV is extensively used. During CV all SVM training samples (in Table 5.8) are trained and validated in order to generalize the memorization accuracy of the SVM classification engine. The main reason for conducting 10-fold CV is to ensure that the SVM classification engine does not overfit the training data.

For the purpose of applying 10-fold CV, the 162 training samples (in Table 5.8) are split into CV training and CV testing sets such that, 70 percent of the total samples (113 samples) from each class are used for CV training and the remaining 30 percent samples (49 samples) from each class are used for CV testing. This iterative procedure is repeated for 100 trials for 10-fold CV, where on each trial the CV training and CV testing data samples are selected randomly.

The Grid Search method proposed by Hsu *et al.* (2003) (in Section 4.2.4) is used for SVM hyperparameter tuning in this research. In the Grid Search method, exponentially growing sequences of parameters (C, γ) are used to identify SVM hyperparameters obtaining the best 10-fold CV accuracy of 87.83 percent (in Section 5.5.3.3). After Grid Search is complete, the optimum SVM hyperplane parameters are found to be: $C = 64$ and $\gamma = 0.001953125$, as shown in Figure 5.28. Thus, using the optimum set of SVM hyperparameters obtained from the Grid Search method, an average SVM training accuracy of 97.60 percent is obtained using the 49 CV testing samples (see Section 5.5.3.3).

6.1.3.1 Discussion of SVM Training Results

In the C-SVM classification model (Hsu *et al.*, 2003) applied in this research, the parameter C is a SVM hyperparameter that defines the trade-off between the training error and complexity of the model (classification engine). In the dual Lagrangian formulation, the parameter C (in equation (4.30)) defines the upper bound of the Lagrange multipliers α_i, hence, it defines the maximal influence the sample can exert on the solution.

For the trained model developed in Figure 5.29, the SVM hyperparameter C affects the training and memorization accuracy of the SVM classification engine. The reason for this is, since there are 10 bounded SVs (BSVs) in the trained model, thus, $\alpha_i = C$ (in equation (4.35)). Due to this, the Grid Search technique selects the parameter C that defines the optimum trade-off between the training error and the complexity of the model, with parameter $C = 64$, signifying that the training data has significant noise. Thus, by using a smaller value of parameter C in the developed model, the results of the SVM classification mapping are smoother with a lower noise consideration. The RBF kernel parameter γ, in the SVM classification engine controls the width of the RBF (Gaussian) kernel. The γ parameter is related to σ, which is defined by the following expression:

$$\gamma = \frac{1}{2\sigma^2} \tag{6.7}$$

where σ^2 is the variance of the resulting Gaussian hypersphere. The optimum value of the SVM hyperparameter γ in equation (6.7) found using Grid Search is computed to be: $\gamma = 0.001953125$. So, the value of σ can be calculated using the following expression:

$$\sigma = \sqrt{\left(\frac{1}{2\gamma}\right)} \tag{6.8}$$

where σ is computed to be $\sigma = 16$ using $\gamma = 0.001953125$. The value of σ for the trained classifier is acceptable, since any value of σ below 0.01 is considered small and any value of σ above 100 is considered large. The reason for this is, as the parameter σ acts as an important hyperparameter during SVM training, small values of σ lead the model close to overfitting the training data, while large values of σ tend to over-smooth the training data. From the statistical learning theory point of view, small σ values lead to a higher VC-dimension, meaning that too many features are used for machine learning which leads to overfitting, while large σ values lead to a lower VC-dimension, signifying that too few features are used to model the classification engine. Thus, the value $\sigma = 16$ is acceptable to model the SVM classification engine using the RBF kernel.

6.1.4 SVM Testing and Validation

The accuracy of SVM testing and validation of is a gauge to evaluate the capability of the developed framework, namely the capability to classify between malignant and benign samples. In this research SVM testing and validation is performed by integrating the LIBSVM v3.0 library (Chang & Lin, 2010) into MATLAB as indicated in Section 5.5.3.4.

The trained model in Figure 5.29 is validated with the 70 testing samples (in Table 5.8) in order to classify previously unseen (untrained) samples. As observed from Figure B.5 in Appendix B, the SVM testing accuracy obtained for an average of 100 trials using 70 testing samples (selected randomly on each trial) is found to be

97.14 percent. In addition, the SVM probability estimates of the tested samples (see Section 5.5.3.4) are obtained with the SVM classification results (class labels). The probability estimates (or scores) can be taken as a measure of confidence during classification of the testing samples, as indicated in Figure 5.31 and Figure 5.32. The experiments performed in this research are presented in Section 6.1.4.1 and discussed in Section 6.1.4.2.

6.1.4.1 SVM Classification Results

The framework developed in this research for the classification of malignant and benign abnormalities (in Figure 5.2) is tested using a Dell XPS 430 Workstation, with a 3.00 GHz Intel Core2 Quad Processor and 8.00 GB of RAM. The time taken for testing one sample approximately takes 4 seconds, which varies based on the configuration of the computer used and the number of samples tested. The following sections present the experiments performed in order to meet the objectives and contributions of this research outlined in Section 1.2 and Section 1.3 respectively.

6.1.4.1.1 Optimum ROI Size Selection

In general it is difficult to determine the size of the neighbourhood or the Region of Interest (ROI) that should be used to extract the relevant GLCM texture features from the abnormal regions (mass lesions and MCCs). If the size of the ROI is too large, small lesions may be missed; while if the ROI size is too small, parts of large lesions may be missed.

The primary contribution of this research as indicated in Section 1.3 is to determine the most suitable ROI (neighbourhood) size in order to perform

optimum texture feature extraction. This specifically addresses the problem of predetermining the ROI size for feature extraction. Thus, in this research, seven common ROI sizes have been evaluated as discussed in Section 5.4.1, namely: 48 × 48 pixels, 64 × 64 pixels, 96 × 96 pixels, 110 × 110 pixels, 128 × 128 pixels, 136 × 136 pixels and 148 × 148 pixels.

Table 6.1: Comparison of classification accuracy using different ROI sizes

No.	ROI Size (in pixels)	Optimum SVM Hyperparameters	Average SVM Accuracy for 100 trials
1.	48 × 48	C = 64, γ = 0.0078125	86.56%
2.	64 × 64	C = 1024, γ = 0.001953125	89.33%
3.	96 × 96	C = 256, γ = 0.001953125	93.87%
4.	110 × 110	C = 32, γ = 0.00390625	94.58%
5.	128 × 128	C = 64, γ = 0.001953125	97.60%
6.	136 × 136	C = 512, γ = 0.0009765625	93.53%
7.	148 × 148	C = 256, γ = 0.00390625	92.76%

Testing the significance of the ROI sizes is performed using 70 testing samples (30 percent of the total ROI samples) with the developed SVM classification engine. The experimental results obtained using different ROI sizes and their tuned SVM hyperparameters, are shown in Table 6.1. As indicated from Table 6.1, the ROI size of 128 × 128 pixels obtains the highest performance of 96.60% in terms of classification between malignant and benign ROIs. Further testing for significance shows that using a ROI size of 128 × 128 pixels results in the lowest number of FPs and FNs (see Table 6.1) as compared to the other six ROI sizes.

In addition, performing analysis on the GT data, the minimum and maximum diameter in pixels of a circle enclosing all malignant and benign abnormalities is

found to be 48 and 130 pixels respectively. Given the above reasons, it is confirmed that a 128 x 128 pixel square ROI (or a 128 pixel circle diameter) is a near optimum to the value that can be used to extract all the abnormal (malignant and benign) regions. All experiments performed from here onwards use a ROI size of 128 × 128 pixels for the purpose of texture feature extraction.

6.1.4.1.2 SVM Testing

In this research, the SVM classification engine is developed using 162 training samples (70 percent of the total ROI samples) for a binary classification problem, where malignant samples are taken as the $positive\ (+ve)$ class and benign samples are taken as the $negative\ (-ve)$ class. Thus, representing the ROI samples as positive and negative instances of a binary classification problem, a confusion matrix can be derived, as indicated in Figure 6.2.

		Actual value		
		p	***n***	**Total**
Prediction outcome	***P'***	True Positive (TP)	False Positive (FP)	P'
	n'	False Negative (FN)	True Negative (TN)	N'
	Total	P	N	

Figure 6.2: Binary classification confusion matrix

Testing the 70 samples indicated in Table 5.8 with the SVM classification engine in Figure 5.29, the resulting confusion matrix obtained with performance indices TP, FP, FN and TN is shown in Figure 6.3.

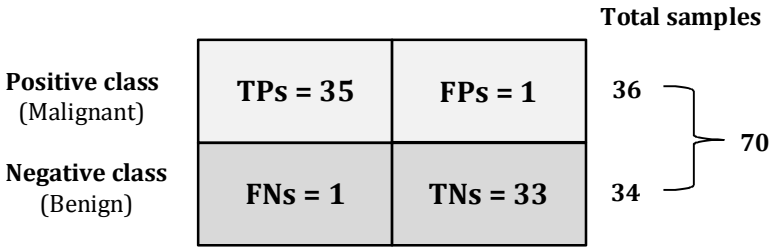


Figure 6.3: Confusion matrix after SVM testing
(malignant is the $+ve$ class and benign is the $-ve$ class)

The SVM testing results obtained in the binary confusion matrix in Figure 6.3 show that, 30 out of the total 31 malignant ($+ve\ class$) samples are classified correctly by the SVM, whereas 38 out of the total 39 benign ($-ve\ class$) samples are correctly classified by the SVM. This indicates that only one sample in both classes is misclassified. Thus, in total 68 out of the 70 tested samples (in Table 5.8) are classified correctly by the SVM, which give a binary classification accuracy of 97.14 percent as indicated in Appendix B in Figure B.5. Using the confusion matrix results in Figure 6.3, the four binary classification performance metrics defined in Table 3.2, namely the TP, FP, TN and FN are computed as shown in Table 6.2.

The *sensitivity* and *specificity* metrics (in equations (3.1) and (3.2)) from the confusion matrix performance metrics are computed to be 0.9710 and 0.9706 respectively, where the minimum and the optimum values of both are 0 and 1 respectively. The classification accuracy is computed using equation (3.3), which is found to be 97.15 percent, where 68 out of the total 70 tested samples are classified correctly by the SVM. Since the sensitivity, specificity and accuracy values are greater than 0.95 (95 percent), thus, the performance of the developed framework is acceptable.

Table 6.2: Binary classification performance metrics using the SVM as the learning machine

Binary Classification Performance Metrics	Equation No.	Value
True Positive (TPs)	-	35
False Positive (FPs)	-	1
False Negatives (FNs)	-	1
True Negatives (TNs)	-	33
Sensitivity	(3.1)	97.10%
Specificity	(3.2)	97.06%
Accuracy	(3.3)	97.14%
True Positive Fraction (TPF)	(3.4)	0.9710
False Positive Fraction (FPF)	(3.5)	0.0290

The True Positive Fraction (TPF) (also known as the sensitivity) and False Positive Fraction (FPF) metrics are calculated using equations (3.4) and (3.5), which are found to be 0.9710 and 0.0290 respectively as shown in Table 6.2. The TPF determines the performance of the SVM classification engine on identifying positive (malignant) samples correctly from all positive samples tested. In contrast, the FPF determines how many incorrect positive results occur among all negative (benign) samples tested.

To visualize binary classification results of the developed framework in Figure 5.1, an ROC curve is plotted using the 70 testing samples (in Table 5.8), as shown in Figure 6.4. Each instance (testing sample) in the binary confusion matrix in Figure 6.3 is represented as one point in the ROC space in Figure 6.4.

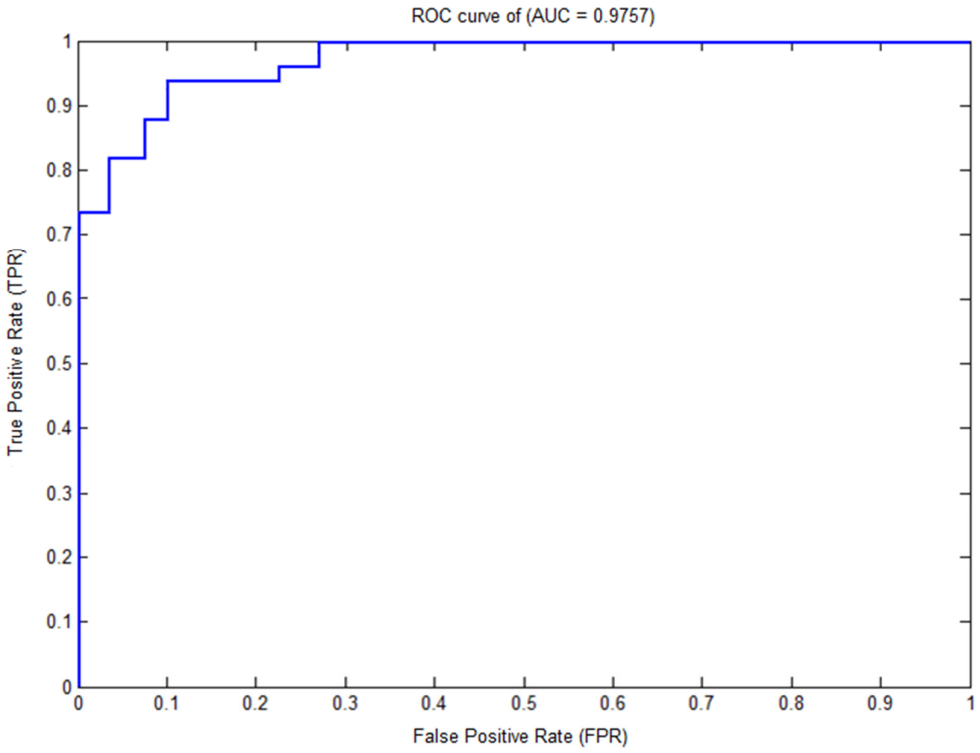


Figure 6.4: ROC curve of SVM classifier for testing with 70 samples (malignant is the $+ve$ class and benign is the $-ve$ class)

The Area Under Curve (AUC) for the ROC curve in Figure 6.4 is found to be $A_z = 0.97574$. The optimum value for the AUC is 1, where ROC curves with $A_z \geq 0.9$ are rated as optimum classification results. As observed from the plot in Figure 6.4, the ROC curve follows close to the left-hand border and then the top border of the ROC space, which indicates that the developed framework produces optimum results for classification between malignant and benign samples.

6.1.4.1.3 False Positive (FP) Reduction Results

The decision-logic system presented in Section 5.5.3.5 and shown in Figure 5.32, reduces the number of FPs for the confusion matrix in Figure 6.3. Each FP instance satisfying the condition ($t < 0.08$) in Figure 5.32 is classified as TN instead of FP, the result of which is shown in the confusion matrix in Figure 6.5.

Positive class (Malignant)	**TPs = 35**	**FPs = 0**
Negative class (Benign)	**FNs = 2**	**TNs = 34**

Figure 6.5: Confusion matrix after implementation of decision-logic system (malignant is the $+ve$ class and benign is the $-ve$ class)

Using the confusion matrix in Figure 6.5, the FPF is calculated using equation (3.5) and is found to be 0, which is an ideal value for the FPF. Using the proposed decision-logic system with a small number of testing samples (70 samples in this case) a FPF of 0 is achievable. However using this decision-logic system, an ideal FPF of 0 cannot be guaranteed unless a larger amount of samples are tested.

One of the limitations in this research concerns the number of mammogram samples acquired for development of the computerized breast cancer detection system. The total number of mammography images obtained from University Malaya Medical Centre (UMMC) is limited due to the fact that UMMC have only recently implemented digital mammography in 2008. Thus, over a course of nearly two years, only a limited number of malignant and benign cases (in Table 5.1) are available from the UMMC in digital format.

6.1.4.2 Discussion of SVM Classification Results

This section summarizes the SVM classification results obtained from experimental testing in Section 6.1.4.1, where four major experiments are performed. All results presented in this section are evaluated on the UMMC and MIAS ROI samples in Table 5.8.

The first experiment presented in Section 6.1.4.1.1 evaluates different ROI sizes in order to determine an optimum ROI size, the experimental results of which are presented in Table 6.1. The optimum ROI size is found to be 128 × 128 pixels with a classification accuracy of 97.60 percent between malignant and benign ROI samples. Since the classification accuracy is greater than 0.95 (95 percent), thus, the performance of the proposed model is acceptable.

The second experiment in Section 6.1.4.1.2 computes the four binary classification performance metrics (TP, TN, FP and FN) using the SVM testing results from the first experiment. The performance metrics are used to plot an ROC curve obtained by testing the 70 samples (in Table 5.8), as shown in Figure 6.4. The ROC curve yields an AUC of $A_z = 0.97574$. Based on a collective comparison of the results obtained from the first and second experiment, the following observations are made:

1. All binary performance metrics in Table 6.2 are greater than 95 percent.
2. The FPF in equation (3.5) is less than 5 percent.
3. The ROC curve follows close to the left-hand border and the top border of the ROC space.
4. The ROC $A_z \geq 0.9$.

These observations indicate that the developed system can classify between malignant and benign ROIs with an average classification accuracy of 97 percent. Since the classification accuracy of the developed system is greater than the baseline of 95 percent, thus it is confirmed that the developed framework shown in in Figure 5.1 produces promising classification results.

The third experiment in Section 6.1.4.1.3 gives attention on reducing the number of FPs obtained from the SVM classification results in Table 6.2. The number of FPs effect the FPF (in equation (3.5)), which can be reduced by applying a decision-logic system (in Figure 5.32) using the probability estimates of the tested samples from the SVM classification results. Applying the decision-logic system confirms that number of FPs and the FPF can be minimized at a low cost. However, since the number of samples in the MIAS and UMMC datasets is less the accuracy of the FPF the reduction algorithm cannot be tested in depth.

6.2 Comparison of Proposed Framework with Other Techniques

In order to estimate the performance of the SVM based model, different machine learning algorithms other than SVM are evaluated. Since, ANNs have similar structure to that of SVMs, thus, they are used in this research comparison with the proposed SVM framework. Traditional and modern ANN based machine learning algorithms namely the Back-Propagation Neural Network (BPNN) and the Online-Sequential Extreme Learning Machine (ELM) presented in Sections 4.3.1 and 4.3.2 respectively are used as the learning machines in the framework in Figure 5.2.

6.2.1 Experimental Results of Compared Techniques

Comparing the developed framework (using SVM) with a traditional and a modern ANN based approach, namely the BPNN (see Section 4.3.1) and the OS-ELM (see Section 4.3.2), provides a better estimate of the memorization and generalization capability of different learning machines.

The BPNN during training uses a different approach in the calculation of the training error, as it minimizes the empirical error, whereas the SVM minimizes the

structural risk. Similar to the BPNN, the OS-ELM is a Single Layer Feed-forward Neural Network (SLFN). Conventional ANN learning algorithms of SLFNs require tuning of network parameters. However, the OS-ELM randomly generates the input weights and the hidden neuron biases of the SLFN and uses them to calculate the output weights without requiring further learning. The OS-ELM implemented in this research is an online variant of the ELM algorithm, applicable for batch learning (Liang *et al*., 2006).

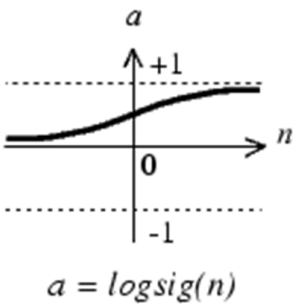


Figure 6.6: Log-sigmoid transfer function

The network architecture of the BPNN implemented in this research consists of 1056 input neurons in the input layer, corresponding to the optimum subset of 1056 texture features (see Tables 5.6 and 5.7). The output layer of BPNN consists of a single neuron, where an output of 0 indicates a benign sample and an output of 1 indicates a malignant sample. In the BPNN, the output of the neurons in the hidden layers is calculated using the log-sigmoid activation function defined in equation (4.46) and shown in Figure 6.6.

The number of samples selected for BPNN training and testing is indicated in Table 5.8. Three parameters need to be determined for the BPNN prior to obtaining a trained model (classifier), which are as follows:

1. Number of hidden layers (h)
2. Number of hidden layer neurons (n)
3. Number of training iterations (e)

In order to statistically determine the optimum parameter values for the BPNN, different combinations of parameter values of n, h and e are iterated for the following ranges: $1 \leq h \leq 20$, $1 \leq n \leq 1000$ and $10 \leq e \leq 2000$. In order to perform 10-fold CV, the 162 training samples (in Table 5.8) are split into CV training and CV testing sets, such that 70 percent of the total samples from each class are used for CV training (113 samples) and the remaining 30 percent samples from each class are used for CV testing (49 samples). This procedure is repeated for 100 trials using 10-fold CV, where on each trial CV training and CV testing samples are selected randomly. The final structure of the BPNN after parameter optimization results in a training accuracy of 93.58 percent (computed using equation (5.21)), where the optimum BPNN parameters determined and used for training are shown in Table 6.3. The BPNN classification results obtained after testing the 70 samples (in Table 5.8) are shown in Table 6.4. The ROC curve obtained from the BPNN classification results for testing with 70 samples is shown in Figure 6.7 with an AUC of $A_z = 0.8235$.

Table 6.3: Optimum parameters for the BPNN modeling

BPNN Parameters	Optimum Value
Number of hidden layers	$h = 3$
Number of hidden layer neurons	$n = [40\ 90\ 30]$ where, n is a matrix specifying the number of hidden neurons in each hidden layer of the BPNN.
Number of training iterations	$e = 120$

The OS-ELM is implemented in this research using the RBF activation function. Using the RBF nodes, the centers and widths of the nodes are randomly generated and fixed, based on this, the output weights are determined by the network.

The network architecture of the OS-ELM implemented in this research consists of 1056 input neurons in the input layer, corresponding to the optimum subset of 1056 texture features (see Section 5.4.3). The output layer of OS-ELM consists of a single neuron, where an output of 0 indicates a benign sample and an output of 1 indicates a malignant sample. In the OS-ELM only one parameter needs to be determined, which is the number of hidden layer neurons n, since the OS-ELM is a SLFN. The method to search for the optimal number of the hidden layer neurons n in the OS-ELM is suggested by (Huang *et al.*, 2004), which indicates that the number of hidden neurons, vary in the range from 20 to 200 as discussed in Section 4.3.2.

Table 6.4: Comparison of the developed framework using different machine learning techniques

Binary Classification Performance Metrics	SVM	BPNN	OS-ELM
TPs	**35**	30	33
FPs	**1**	6	3
FNs	**1**	5	4
TNs	**33**	29	30
Sensitivity	**97.10%**	85.71%	89.19%
Specificity	**97.06%**	82.86%	88.24%
Accuracy	**97.14%**	84.29%	90.00%
TPF	**0.9710**	0.8571	0.8919
FPF	**0.0290**	0.1429	0.1081
AUC (A_z)	**0.97574**	0.8235	0.8971

The optimal value of n is determined based on the classification performance of the OS-ELM, which is the training accuracy (equation (5.21)). Since the number of neurons in the input layer of the OS-ELM is large i.e., 1056, thus, for modeling purposes, the range of n is selected as $10 < n \leq 1000$, where the size of n is incremented by a value of 10 on each iteration.

The final architecture of the OS-ELM after parameter optimization results in a training accuracy (equation (5.21)) of 96.28 percent, where the optimal size of the hidden layer neurons computed be $n = 160$. The OS-ELM classification results obtained after testing the 70 samples (in Table 5.8) are shown in Table 6.4. The ROC curve obtained from the OS-ELM classification results for testing with 70 samples is shown in Figure 6.8 with an AUC of $A_z = 0.8971$.

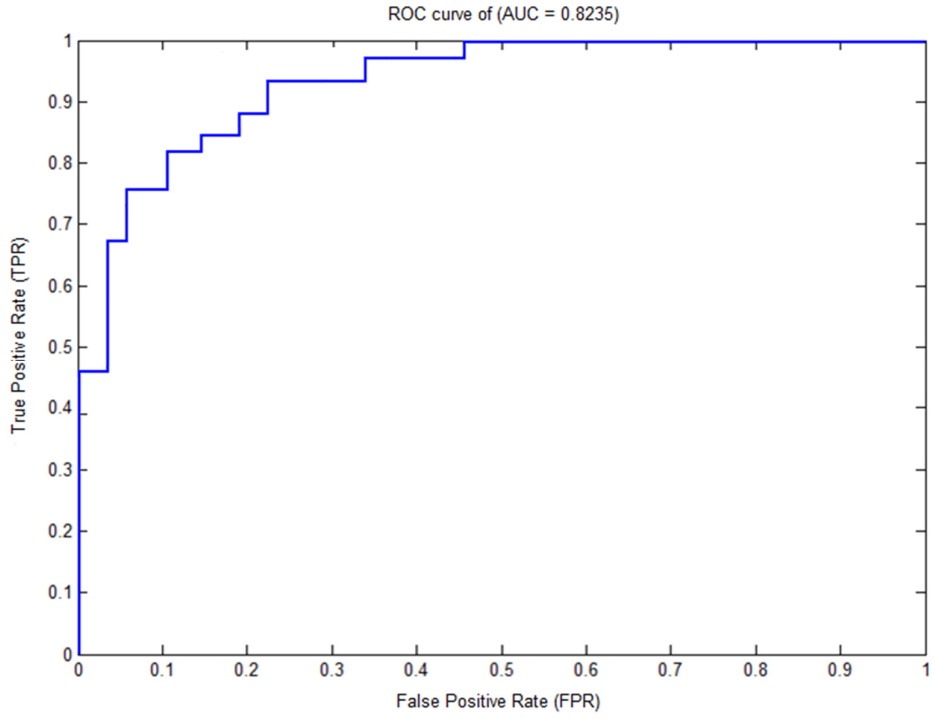


Figure 6.7: ROC curve of BPNN classifier for testing with 70 samples (malignant is the $+ve$ class and benign is the $-ve$ class)

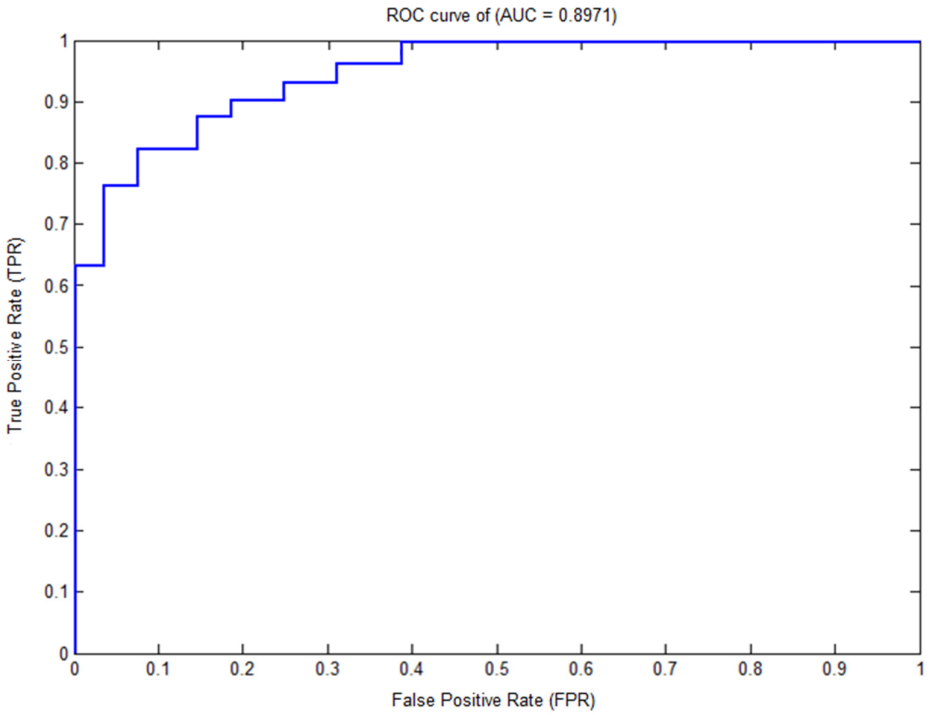


Figure 6.8: ROC curve of OS-ELM classifier for testing with 70 samples (malignant is the $+ve$ class and benign is the $-ve$ class)

Parameter optimization for the BPNN and the OS-ELM in this research is performed using 10-fold CV, which is similar to the case of the SVM. The reason for using CV is that, since the number of training samples can be divided further into subsets, CV ensures that the trained model (classification engine) does not overfit the training data.

6.2.2 Discussion of Compared Models

Both SVMs and ANNs are considered as black-box modeling techniques. Although both algorithms share the same structure, but the learning methods for both algorithms are completely different. ANNs try to minimize the training error, whereas SVMs reduce capacity using the SRM principle.

Comparison results of BPNN and OS-ELM in contrast to the SVM based model are tabulated in Table 6.4, which are obtained for testing the 70 samples from the local dataset (UMMC and MIAS). The experimental results in Table 6.4 show that the SVM based approach outperforms the BPNN and the OS-ELM with respect to the overall classification accuracy. This is because the optimum results for binary classification are obtained by the SVM based model, where parameters: sensitivity, specificity, TPF, FPF and A_z are in optimum ranges.

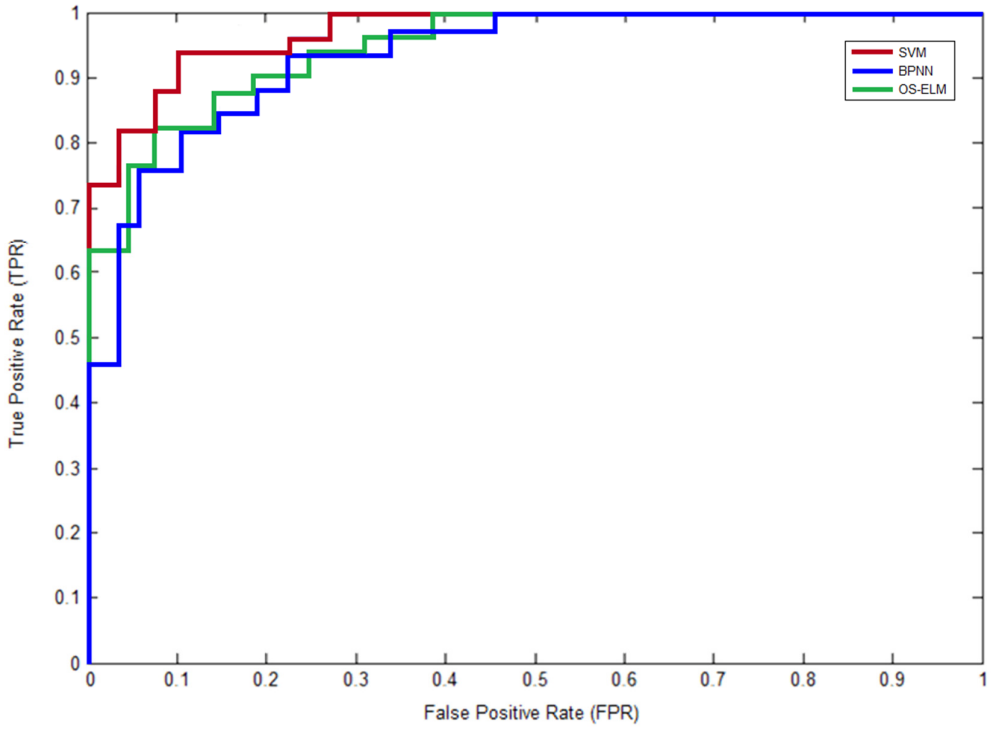


Figure 6.9: ROC curves indicating the performance of the compared machine learning techniques

To further investigate the accuracy of the compared machine learning models, the ROC curves of all three models are computed using statistics from Table 6.4, as shown in Figure 6.9. As observed from Figure 6.9, the SVM has the highest AUC ($A_z = 0.97574$), followed by the OS-ELM ($A_z = 0.8971$) and the BPNN

($A_z = 0.8235$). The curve for the SVM follows the closest to the left-hand border and then the top border of the ROC space this indicates that the SVM has better classification results compared to the other techniques.

As observed from Table 6.4, the BPNN has the lowest performance in terms of the classification accuracy out of all compared models. Since the BPNN used in this research has a training accuracy of 93.58 percent, which is higher than the generalization (testing) accuracy of 84.29 percent, this indicates that the BPNN has a lower generalization as compared to the OS-ELM and the SVM. The main reason for the low generalization of the BPNN is due to the cause of excessive training, i.e. overfitting.

During BPNN training, the goal is to obtain a *global optimum* solution. However, in a BPNN, to get the overall minimum answer of the error function, the network extrema corrects itself slowly along the local improved way and eventually ends up obtaining the local optimization answers only, which generally occurs due to excessive training (overfitting). The reason for this is, since the BP algorithm is based on the gradient descent approach, the network descends slowly with a low learning speed and when a flat section (roof) appears for a long time the algorithm ends the training at that instance, which results in locally optimized answers.

Another reason for the low generalization of the BPNN is due to noise in the digital mammography data (features). The low generalization of the BPNN does not mean that it is not a good tool for pattern classification, but given the reasons above, it is not a considered a suitable tool to be evaluated with the mammography datasets (in Table 5.1) acquired in this research.

In terms of classification performance, from Table 6.4 it is observed that the OS-ELM ranks second after the SVM, whereby the BPNN ranks the last. The reason for the better classification accuracy of the OS-ELM compared to the BPNN is that, since the OS-ELM iteratively fine tunes the network's input weights and biases using finite samples of the training data, this yields a higher generalization for the OS-ELM. The RBF transfer function applied in the OS-ELM technique randomly initializes hidden neuron parameters such as the: input weight vectors, neuron biases for additive hidden neurons, centers and impact factors for RBF hidden neurons, and iteratively computes the output weight vectors.

During experimental testing of the OS-ELM technique it is observed that, if the order of the training samples is switched or changed, the training accuracy of the OS-ELM also changes significantly. In order to obtain an average estimate of the memorization performance of the OS-ELM, the training accuracy of the OS-ELM is computed using an average of 100 trials where on each trial training samples are selected randomly. It is observed from the OS-ELM that, with the increase in the number of input layer neurons, the OS-ELM achieves a better performance, while remaining stable for a wide range of input neuron sizes.

There are a few reasons which constitute to the low performance of the OS-ELM compared to the SVM. The first reason being that, the assignment of the initial weights in the OS-ELM is arbitrary, which effects generalization performance of network. As the proper selection of input weights and hidden bias values contributes to the generalization capability of the trained model (classification engine), the initialization of arbitrary weights decreases the generalization performance of the OS-ELM.

The second reason being that, the value of γ parameter in the RBF activation function of the OS-ELM is set as a constant value of 1, as discussed by Huang *et al*., (2006b) and Liang *et al*. (2003). As the parameter γ controls the width of the RBF function in the OS-ELM, thus, it is suggested to be selected in within the range of 0 to 1. If the value of γ is increased to $\gamma \geq 1$, the generalization performance for unseen data will decrease.

More importantly, the value of γ cannot be fixed to a constant value, since the width of the Gaussian function depends upon the data samples to be classified and also the amount of noise present in the data. Since there is no evidence or literature on the ELM on how to tune the parameter γ for the RBF activation function, using the default parameter $\gamma = 1$ as suggested by Huang *et al*., (2006b) and Liang *et al*. (2003), the OS-ELM produces lower generalization performance compared to the SVM. The OS-ELM does suffer from a few drawbacks, which are as follows:

(a) For achieving good generalization results with the OS-ELM, the number of hidden layer neurons (n) must be chosen larger than standard ANN algorithms, (such as the BPNN). This is because the neuron weights and biases are not learned from the training data.

(b) Multi-layer ANNs (such as the BPNN used in this research) if trained properly, can possibly achieve similar and even better results comparable to the OS-ELM, a SLFN.

(c) The solution provided by the ELM and the OS-ELM is not always so smooth, and mostly shows some ripple.

The only notable advantage of the OS-ELM over the SVM is its faster training process, with the increase in the chunk (data) size. It is known that using the RBF (Gaussian) as the activation function, SVMs suffer from tedious parameter tuning. However, the OS-ELM with a single parameter (n) to be tuned uses its arbitrary assignment of initial random weights, which requires it to search for the optimal size of hidden layer neurons (n). This requires the OS-ELM to execute many times in order to get an average estimate, which loses its edge over the SVM.

The experimental results presented in Section 6.2.1 indicate that using the SVM for classification of malignant and benign abnormalities from digital mammography data has shown to be very promising. In this research, SVMs have the a few notable advantages as compared to ANNs, which are as follows:

- SVMs have non-linear dividing hypersurfaces that give them high discrimination.
- They provide good generalization ability for unseen data classification.
- They determine the optimal network structure (such as the hidden layers and hidden layer neurons) themself, without requiring to fine tune any external parameters.

In contrast to the advantages of SVMs over ANNs, there are some drawbacks of SVMs. However, these drawbacks are restricted due to practical aspects concerning memory limitation and real-time training of SVMs. The drawbacks of SVMs are as follows:

(a) The quadratic programming (QP) optimization problem arising in SVMs is not easy to solve. Since the number of Lagrange multipliers is equal to the number of training samples, the training process is relatively slow. Even with the use of the Sequential Minimal Optimization (SMO), real-time training is not possible for large datasets.

(b) The second drawback of SVMs is the requirement of storage capacity for the trained model (classification engine). Support vectors (SVs) in the trained model represent important features distinguishing the training samples between the two classes (malignant and benign). When the optimization problem has a low separability in the space used, the number of SVs increases. SVs have to be stored in a model file. This puts limitations on the implementation of SVM for devices with limited storage capacity.

Given all these aspects, the experimental results presented in Table 6.4 and Figure 6.9 shows that the SVM provides a better classification performance compared to traditional and modern ANN based approaches. Thus, SVMs are considered as a superior machine learning technique when the requirement is to solve classification problems with noisy data.

6.3 Summary

This chapter presented the experimental results of the developed system in Chapter 5. Section 6.1 presented and discussed the SVM training results relative to the memorization and learning of the binary SVM classifier. Section 6.1 also presented and discussed the SVM testing and validation results for unseen samples. In order to perform a comparative research, Section 6.2 presented the experimental results obtained after evaluating the developed framework using

different machine learning algorithms other than the SVM. The experimental results of the compared machine learning models are discussed in the last part of Chapter 6.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.0 Overview

This chapter concludes and summarizes the research contributions made. The achievements and objectives of the research with respect to the experimental results obtained are highlighted along with the key findings and significance of the research. This chapter also discusses the impact and significance of the developed system to radiologists and hospitals for mammography screening and interpretation. Radiologists and clinicians will benefit from the developed system as it will assist them in their diagnosis by acting as second readers.

7.1 Benefits of the Developed System

Digital mammography leads itself well to computerized detection of breast cancer, where computer-aided methods based on image processing and machine learning algorithms enable computers to identify suspicious areas of the breast that can be mass lesions, MCCs or other signs of breast cancer.

In this research, an approach towards image processing and machine learning are applied to develop a breast cancer detection system for classification between malignant and benign abnormalities in digital mammograms, as shown in Figures 1.4 and 5.1. The modeling and development of the proposed system is presented in Chapter 5. Firstly, the acquired digital mammography images in Table 5.1 are preprocessed and segmented. Next, texture features are extracted from the

segmented mammogram images, where the optimal subset of features are selected and classified using SVMs and ANNs. The experimental results presented in Section 6.2.2 (in Table 6.4) indicate that SVMs provide better classification performance compared to traditional and modern ANNs.

Computerized breast cancer detection has provided a huge benefit in hospitals, which are constantly looking for expert radiologists, and would have sensible effects on medical and ethical grounds. Computer-aided methods have the potential to increase the diagnostic accuracy by reducing the FPF, while increasing the positive predictive values (PPVs) of mammographic abnormalities as discussed in Chapter 3. The benefits obtained of using the developed system for breast cancer detection are as follows:

1. This system will aid radiologist's clinicians in the mammography screening and interpretation process by acting as a second reader after the radiologists.

2. This system will substantially reduce the number of false positives (FPs), (see Section 3.3.6), which will eliminate the need of performing unnecessary biopsies and save cost.

3. This system will reduce patient examination time by inspecting mammograms and reporting the findings within a few seconds.

The weakest link in breast cancer detection has always been the radiologists, since it is the radiologists who must find masses and MCCs. In cases where radiologists have difficulties identifying between cancerous and non-cancerous abnormalities, they can refer to this system for a second opinion as it is often difficult to distinguish between malignant from benign abnormalities due to their similar nature and visual features.

7.2 Contribution and Significance of Research

The detection of breast cancer in digital mammography applications using machine learning approaches requires *feature/heuristic computation* (Woods and Bowyer, 1996) from the Region of Interest (ROI), namely, the abnormal region. As discussed in Section 1.3 previously, it is difficult to determine the size of the neighbourhood (pixels) or the ROI that should be used to calculate the relevant features from the abnormal regions (masses and/or MCCs). If the size of the ROI is too large, small masses and/or MCCs may be missed, while if the size of the ROI is too small, parts of large masses and/or MCCs may be missed. This poses a challenging task in the computerized detection of breast cancer. Thus, the primary contribution of this research as outlined in Section 1.3 is:

- To determine the most suitable ROI (neighbourhood) size of mass lesions and MCCs for the purpose of feature computation (extraction).

The experimental results presented in Section 6.1.4.1.1 of this thesis contribute to the problem of determining the optimum ROI size using the digital mammographic data acquired in this research (in Table 5.1). Table 6.1 shows the experimental results of evaluating different ROI sizes. It is observed from Table 6.1, that the ROI

with a size of 128 × 128 pixels achieves optimum results, with a classification accuracy of 97.14 percent between malignant and benign samples. The experimental results in Table 6.2 show that promising classification results can be obtained by selecting the optimum ROI size for the purpose of feature computation. The secondary contribution of this research outlined in Section 1.3 is:

- To demonstrate that advanced machine learning techniques, namely, Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) can effectively solve pattern classification problems.

The secondary contribution of this research provides the basis for conducting a comparative analysis between different machine learning technologies, such as SVMs and ANNs. Since SVMs and ANNs are both learning machines, which share the same structure but utilize different learning methods, two ANN approaches, namely, the Back-propagation Neural Network (BPNN) (traditional approach) and the Online-Sequential Extreme Learning Machine (OS-ELM) (modern approach) are evaluated in this research for comparison with SVM.

The experimental results presented in Section 6.2.2 compares the three learning machines: the BPNN, the OS-ELM and the SVM. As observed by the experimental results in Table 6.4 and Figure 6.9, the SVM outperforms the BPNN and the OS-ELM techniques in terms of classification performance. This indicates that the SVM has a better generalization capability for the classification of malignant and benign patterns as compared to traditional and modern ANN based techniques.

SVMs have a considerable advantage over ANNs, as they provide the use of soft margins for the purpose of classification (see Section 4.2.4), thus allowing improvement in the generalization performance of the developed system. In this research, the observed advantages of SVMs over ANNs are as follows:

- SVMs have non-linear dividing hypersurfaces that give them high discrimination capability, which is not the case with ANNs.
- SVMs provide good generalization ability for unseen data classification, as they determine the optimal network structure themself, which is not the case with ANNs.
- With the introduction of the SVM, the developed system is able to control the balance between the sensitivity and the specificity, giving it more flexibility.

The SVM classification engine developed has a good memorization capability. As indicated from the experimental results in Section 6.1.3, the training accuracy of the SVM (equation (5.21)) averaged over 100 trials is 97.60 percent, whereas the 10-fold CV accuracy is 87.83 percent. A training accuracy of greater than 95 percent indicates that the memorization and learning capability of the learning machine is notably good, even with the presence of noisy data. The 10-fold CV accuracy is taken as a measure to ensure the trained model (classification engine) does not overfit the training data, where CV accuracy in between 80 to 95 percent typically indicates good memorization capability with no overfitting. The two main reasons contributing to the good training accuracy of the SVM classification engine are:

1. Selection of the optimal subset of texture features for the learning machine (SVM), as presented in Section 5.4.3.
2. Fine tuning of the SVM hyperplane parameters (C, γ) using the v-fold CV approach, as presented in Section 5.5.3.1.

The experimental results obtained from testing the SVM classification engine with unseen samples from the local (UMMC and MIAS) mammography datasets provide a classification accuracy of 97.14 percent (in Table 6.4) for 70 testing samples. Classification accuracy greater than 95 percent indicates promising results for unseen data classification, for any learning machine.

It is known that the performance any learning machine can be problem dependent, since the performance is based on a few factors such as: the experimental datasets used, the optimum subset features selected for modeling and the method in which the data samples are split between the training and testing sets. It is worth noticing that SVMs have indicated lower classification performance compared to ANN techniques, as reported by Osareh *et al.* (2002). Thus, the suitability of a learning machine for a pattern classification task is data dependent. Since the local datasets used in this research contain a noisy data, the SVM is found to be the most suitable technique for classifying between malignant and benign patterns.

7.3 Achievement of Research Objectives

Digital mammography is a relatively new technique for the early detection of breast cancer. It is based on accumulated density of tissues, i.e. to detect shadows. This is the reason as to why mammography has been considered as an efficient

tool for the detection of masses and MCCs (Khuzi *et al.*, 2009), (Verma & Zakos, 2001), (Jiang *et al.*, 1999).

As discussed in Section 1.2, the goal of this research is to increase the diagnostic accuracy of image processing and machine learning techniques for optimum classification between malignant and benign abnormalities as well as to the reproducibility of mammographic interpretation. In order to achieve this goal, the research objectives outlined in Section 1.2 have been obtained, which are discussed as follows:

1. The decision-logic system presented in Section 5.5.3.5 has shown to reduce the number of false positives (FPs) (see Section 3.3.6). Moreover, from the experimental results in Table 6.4 it is observed that SVMs are good are reducing FPs. Thus, the techniques and algorithms implemented in this research are capable of reducing the number of misclassified malignant cancers (FPs), which complies with the third research objective of the research.

2. The data acquired in this research is collected from different sources. The acquired data is classified into two types, namely the *local dataset* and the *external dataset*. The local dataset is a collection of digital mammography images acquired from the University of Malaya Medical Centre (UMMC) patient records. The external dataset is a well-known published image database of 322 digital mammograms from the Mammographic Image Analysis Society (MIAS) (in Section 5.2.1). Since two different datasets are evaluated in this research, this complies with the fourth research objective.

3. To perform a comparative research, the experimental results obtained from different learning machines are presented and compared in Chapter 6. The proposed learning machine, the SVM, is compared to modern and traditional ANN based techniques, namely the BPNN and the OS-ELM respectively, as discussed in Section 6.2. The experimental results obtained in Section 6.2.2 indicate that optimum performance for the classification between benign and malignant patterns is obtained by the SVM technique. This indicates the promising results of the SVM. This complies with the fifth research objective.

7.4 Impact and Significance to Radiologists

The framework developed for the computerized breast cancer detection system in Figure 5.1, can be implemented as an intelligent classification system to assist radiologists in their diagnosis by acting as a second reader. The framework shown in Figure 7.1 is termed as an *Intelligent Classification System*, which can be implemented using the framework developed in this research. It is envisaged that this system will aid radiologists in their interpretation of malignant and benign abnormalities.

The intelligent classification system features a graphical user interface (GUI), which allows radiologists to use the developed computerized breast cancer detection system as a user-friendly software application. The intelligent classification system can have up to three inputs (depending upon the nature of diagnosis), and one output. Out of the three inputs, two inputs are compulsory for diagnosis, which are as follows:

1. Digital mammogram image (image to diagnose/interpret).
2. Location of (x, y) co-ordinates of the center of the ROI (malignant and benign abnormalities) to diagnose.

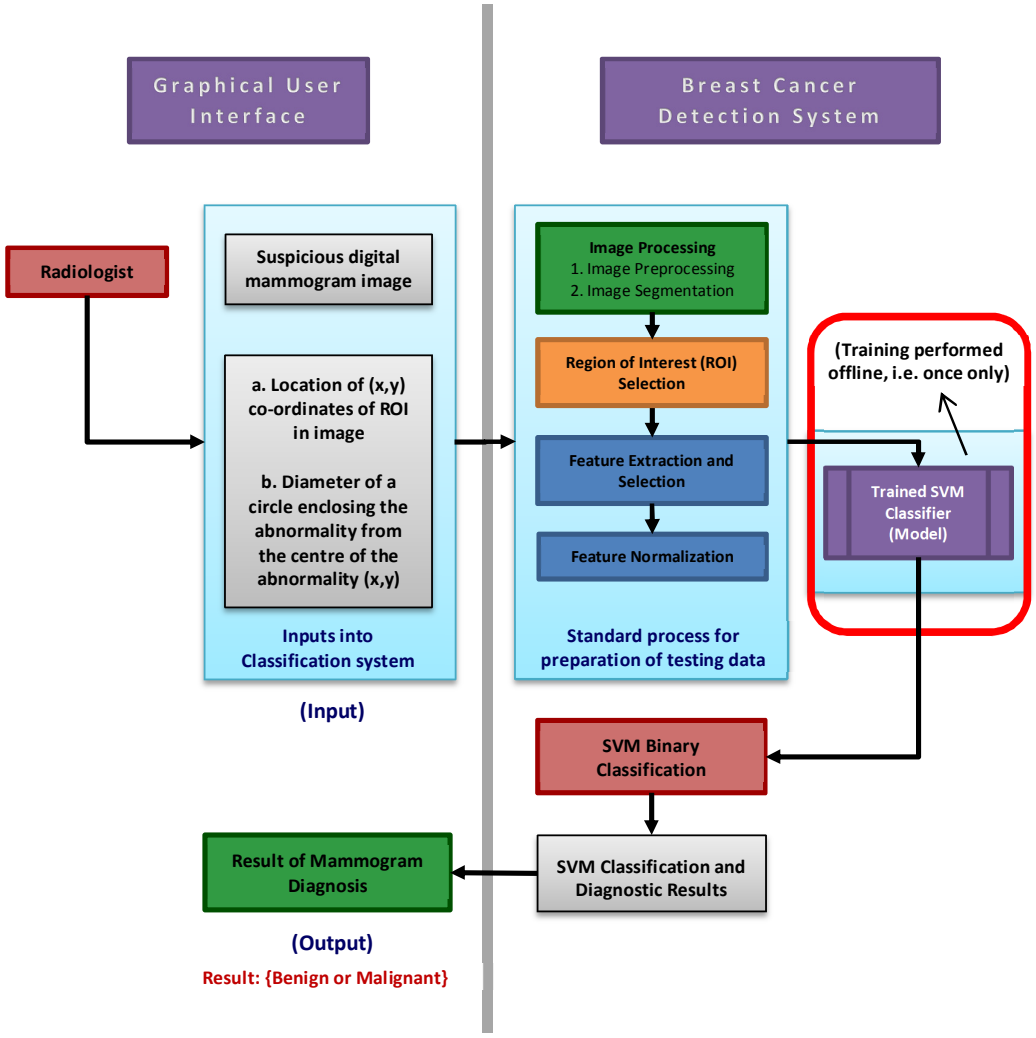


Figure 7.1: Intelligent classification system

The third input into the intelligent classification system can be taken as the diameter of a circle enclosing the abnormality from the centre of the abnormality, which is the location in (x, y) co-ordinates; the *Ground Truth* (GT) data (in Section 5.2.1). Since the value of the diameter optimally determines the size of the ROI (neighborhood), the value of the diameter can be set as a variable parameter in the

GUI system, so it can be adjusted by radiologists based on their findings. Using the ROI size as a variable parameter for the radiologists to specify will yield better results with lesser false positives (FPs) (see Section 3.3.6). The output of the intelligent classification system will indicate if the input mammogram is either a benign or malignant abnormality.

Using the proposed intelligent classification system radiologists can incorporate the output from the computer into their decision. Several recent studies demonstrate that computer-aided detection improves radiologists' ability in differentiating malignant abnormalities from benign ones (Giger, 1999), (Jiang *et al*., 1996), (Jiang *et al*, 1999), (Wu *et al*, 1993), (Huo *et al*., 2000), (D'Orsi *et al*., 1992), (Baker *et al*, 1996), (Chan *et al*, 1999).

The framework developed in Figure 5.1 shows encouraging results as indicated in Chapter 6, however, it also has a few limitations. These limitations can be eliminated by implementing the future work suggested in the following section.

7.5 Future Expansion and Recommendations

Although the computerized breast cancer detection system can achieve a classification accuracy of 97 percent as indicated in Chapter 6, however, this does not guarantee it will obtain good and similar results on other mammography datasets, especially those datasets which have not been tested in this research. Thus, research can be further continued on to improve the performance of the system and validating it by testing with larger digital mammography datasets such as the Digital Database for Screening Mammography (DDSM) (Heath *et al*., 2001).

This researcher strongly believes that the computerized breast cancer detection system developed in this research will contribute significant improvements in mammographic interpretation of cancers. The following sections provide suggestions and recommendations on future work that can be performed in order to enhance the performance of the system and cater for untested mammography datasets.

7.5.1 SVM Parameter Tuning using Genetic Algorithm

For any classification task, the performance of a learning machine will decrease if the modelling (training) parameters are not selected properly. The modelling parameters of a learning machine need to be fine-tuned (optimized) to obtain an optimum balance between the generalization and memorization of the trained model. Lagrangian parameter selection in the case of the SVM is complex in nature, as it is difficult to solve by conventional optimization techniques (see Section 4.2.3.1).

A difficulty of using the SVM is the selection of parameter C and the kernel parameter γ (Gamma) in the RBF (Gaussian) kernel (in equation (4.44)). Even with the use of v-fold CV and the Grid-Search method (Hsu *et al.*, 2003), an optimal solution might not be achieved. The optimum values of the hyperparameters (C, γ) in the SVM need to be found, so they can minimize the expectation of testing and validation error, that needs to adapt to multiple parameters values at the same time.

Genetic Algorithm (GA) with characteristics of high efficiency and global optimization has been widely applied in many applications to solve optimization

problems (Anastasio *et al*., 1998). So, it is suggested that to solve the Dual Lagrangian Optimization (DLO) problem in SVMs, the GA can be applied to optimize the SVM hyperplane parameters (C, γ), This idea proposes a hybrid combination of SVM-GA such that, the GA chromosomes will represent solutions as the (C, γ) parameters and the GA fitness function will evaluate the accuracy of the solutions. The fittest (best) solutions obtained after the iterative GA process stops will be the optimum (C, γ) parameters for the SVM. The hybrid SVM-GA approach suggested here will avoid the local optimum in finding the maximum Lagrangian.

7.5.2 Implementation of Multi-scale RBF Kernel

The RBF (Gaussian) kernel (in equation (4.44)) is one of the well-known Mercer's kernels (Mercer, 1909) for SVMs, which has been widely used in many classification tasks, as the case of this research. The RBF kernel uses the Euclidean distance between two points in the original space to find the correlation in the augmented space. The points very close to each other are strongly correlated in the augmented space, whereas the points far apart are uncorrelated in the augmented space. There is only one parameter to adjust the width of the RBF kernel σ (sigma), which is not powerful enough to cater for complex classification tasks.

The number of features used for machine learning modeling in this research is 1056 (see Tables 5.6 and 5.7), which is large. In order to achieve a better kernel for SVMs using a large number of features, one possible way is to adjust the velocity of decrement in each range of the Euclidean distance between the two points. The multi-scale kernel obtained using this method will maintain the characteristics of a RBF kernel. To implement this multi-scale kernel, a combination of RBF kernels at different scales is suggested as the future work for this research. In order to

proceed, at first, a linear combination of the RBF must be satisfied to be Mercer's kernel. Theoretically, using a multi-scale RBF kernel, the performance and classification accuracy of SVM will improve.

7.5.3 Evaluating Other Texture Approaches

This study uses GLCM texture descriptors for feature computation, as shown in Table 5.7. The reason for using GLCMs in this research is due to the recent success of GLCMs in digital mammography applications, as indicated by the literature review in Table 3.3. However, this does not mean that amongst the texture based techniques, GLCMs can only provide good feature computation results. Thus, other texture based techniques can be evaluated in this research to perform a comparative study. The following texture analysis techniques have gained recent success in pattern classification problems and can be considered:

(a) Spatial Gray Level Dependence Method (SGLDM)

(b) Gray Level Run Length Matrix (GLRLM)

(c) Gray Level Difference Method (GLDM)

A comparative study evaluating other texture feature estimation approaches will benefit this research, with a possibility in further improving the performance and accuracy of the developed system.

7.6 Conclusion

In conclusion, this research has shown encouraging results and a performance that matches human intelligence for classifying between malignant and benign abnormalities in digital mammograms. The goal of this research has focused on

increasing the diagnostic accuracy of computer-aided detection methods used in breast cancer detection.

The experimental results presented in Chapter 6 of this thesis, highlight the significance and key contributions of this research which, the computerized breast cancer detection system developed in this research has an average classification accuracy of 95 percent for the SVM-based model. The system developed in this research has a few notable advantages to radiologists. Firstly, this system will aid clinical radiologists in the mammographic interpretation process by acting as a second reader after the radiologists. Secondly, this system will reduce the number of false positives (FPs), which will eliminate the need of performing unnecessary biopsies and save costs. Lastly, this system will reduce patient examination time by inspecting mammograms and reporting the findings within a few seconds.

As usually happens in an area of research, many approaches can be used and developed, given the appropriate amount of time and effort. It is strongly recommended that the future work suggested in Section 7.5 should be investigated. The proper application and use of the developed system in this research will be appreciated by radiologists in Malaysia. With the remarkable pattern classification capability of SVMs, as shown in the experimental results in Chapter 6, it is desired that more SVM-based applications should be developed for the improvement and quality of health care systems.

REFERENCES

Adams, R. Bischof, L. 1994, *Seeded Region Growing, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 6,* pp. 641–647.

Adler, D.D. and Helvie M.A. 1992, Mammographic biopsy recommendations, *Curr. Opin. Radiol.*, vol. 4, pp. 123–129.

Alexander, F.E. Anderson, T.J. Brown, H.K. Forrest, A.P. Hepburn, W. Kirkpatrick, A.E. Muir, B.B. Prescott, R.J. Smith, A. 1999. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening, *Lancet*, vol. 353, pp. 1903-1908.

American Cancer Society, 2003a. *Breast Cancer: Facts and Figures 2003-2004*, Atlanta, GA: American Cancer Society.

American Cancer Society, 2003b. *Cancer Prevention and Early Detection Facts and Figures 2003-2004*, Atlanta, GA: American Cancer Society.

American Cancer Society, 2009. *Breast Cancer: Facts and Figures 2009-2010*, Atlanta, GA: American Cancer Society.

Anastasio, M.A. Yoshida, H. Nagel, R. Nishikawa, R.M. Doi, K. 1998, A genetic algorithm-based method for optimizing the performance of a computer-aided diagnosis scheme for detection of clustered microcalcifications in mammograms, *Medical Physics*, vol. 25, no. 9, pp. 1613–1620.

Anttinen, I. Pamilo, M. Soiva, M. and Roiha, M. 1993, Double reading of mammography screening films: one radiologist or two? *Clinical Radiology*, vol. 48, pp. 414-421.

Aylward, S.R.Hemminger, B.M. Pisano, E.D. Johnston, R.E. 1998, Mixture modeling for digital mammogram display and analysis. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography*, pp. 305–312. Kluwer, Dordrecht.

Baines, C.J. 1992, Breast self-examination. Cancer, vol. 69, pp. 1942–1946.

Bajger, M. Ma, F. and Bottema, M.J. 2005, *Minimum Spanning Trees and Active Contours for Identification of the Pectoral Muscle in Screening Mammograms*, in Proc. of Digital Image Computing: Techniques and Applications, Dec. 2005, pp. 323–329.

Baker, J.A. Kornguth, P.J. Lo, J.Y. and Floyd Jr., C.E. 1996, Artificial neural network: improving the quality of breast biopsy recommendations, *Radiology*, vol. 198, pp. 131–136.

Ballard, D.H. and Brown, C.M. 1982, *Computer Vision.* Englewood Cliffs, New Jersey: Prentice-Hall.

Bassett, L. W. Jackson, V. P. Jahan, R. Fu Y. S., and Gold, R. H. 1997, *Diagnosis of Diseases of the Breast*, W. B. Saunders Company.

Bazzani, A. Bollini, D. Campanini, R. Riccardi, A. Bevilacqua, A. Lanconelli, N. Romani, D. 2001, Automatic detection of clustered microcalcifications using a combined method and an SVM classifier. In Yaffe, M.J. editor, *Digital Mammography 2000*, in Proc. of the 5th International Workshop on Digital Mammography, Madison, Medical Physics Publishing, pp. 161–167.

Beam, C.A. Sullivan, D.C. 1994, What are the issues in the double reading of mammograms?, *Radiology*, vol. 193, no. 2, pp. 582.

Beam, C.A. Layde, P.M. Sullivan, D.C. 1996, Variability in the interpretation of screening mammograms by u.s. radiologists, *Arch Intern Med*, vol. 156, pp. 209–213.

Behroozmand, R. Almasganj, F. 2005. *Comparison of Neural Networks and Support Vector Machines Applied to Optimized Features Extracted from Patients' Speech Signal for Classification of Vocal Fold Inflammation*", in Proc. of the 5th IEEE Symposium on Signal Processing and Information Technology, pp. 844–849.

Beichel, R. Sonka, M. 2006. Computer vision approaches to medical image analysis. *Lecture Notes in Computer Science*, vol. 4241, Springer, Berlin.

Bellotti, R. De Carlo, F. Tangaro, S. Gargano, G. Maggipinto, G. Castellano, M. Massafra, R. Cascio, D. Fauci, F. Magro, R. Raso, G. Lauria, A. Forni, G. Bagnasco, S. Cerello, P. Zanon, E. Cheran, S.C. Lopez Torres, E. Bottigli, U. Masala, G.L. Oliva, P. Retico, A. Fantacci, ME. Cataldo, R. De Mitri, I. De Nunzio, G. 2006, A completely automated CAD system for mass detection in a large mammographic database, *Med. Phys.*, vol. 33, no. 8, August, pp. 3066-3075.

BI-RADS, Breast Imaging Reporting and Data System, American College of Radiology, 2010. [Online]. Website available at: http://www.acr.org/SecondaryMainMenuCategories/quality_safety/BIRADSAtlas/BIRADSAtlasexcerptedtext/BIRADSMammographyFourthEdition.aspx

Bird, R.E. Wallace, T.W. Yankaskas, B.C. 1992, Analysis of cancers missed at screening mammography, *Radiology*, vol. 184, pp. 613–617.

Bishop, C. 1995, *Neural Networks for Pattern Recognition*: Oxford: Oxford University Press.

Bjurstam, N. Björneld, L. Duffy, S.W. Smith, T.C. Cahlin, E. Erikson,O. Lingaas, H. Mattsson, J. Persson, S. Rudenstam, C.M. Säwe-Söderberg, J. 1997. The Gothenburg Breast Cancer Screening Trial: preliminary results on breast cancer mortality for women aged 39-49, *J Natl Cancer Inst Monogr*, vol. 22, pp. 53-55.

Blot, L. Zwiggelaar, R. 2000a. *Extracting background texture in mammographic images: a co-occurrence matrices based approach*, in Proc. of the 5th International Workshop on Digital Mammography, Toronto, pp. 142–148.

Blot, L. Zwiggelaar, R. 2001. *Background texture extraction for the classification of mammographic parenchymal patterns*, in Proc. of Medical Image Understanding and Analysis, pp. 145–148.

Blot, L. Zwiggelaar, R. Boggis, C.R.M. 2000b. Enhancement of abnormal structures in mammographic images, in Proc. *of Medical Image Understanding and Analysis*, pp. 125–128.

Blot, L. Davis, A. Holubinka, M. Martì, R. Zwiggelaar, R. 2002. Automated quality assurance applied to mammographic imaging, *EURASIP Journal Applied Signal Processing*, vol. 2002, no. 1, pp. 736–745.

Boser, B. Guyon, I. Vapnik, V. 1992. A training algorithm for optimal margin classifiers, in Proc. of the 5th Annual ACM Workshop on Computation Learning Theory, pp. 144–152, ACM Press, Pittsburgh.

Bottema, M.J. Slavotinek, J.P. 1998, Detection of subtle microcalcifications in digital mammograms. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography 1998*, pp. 209–212. Kluwer, Dordrecht.

Bottema, M.J. Slavotinek, J.P. 2001, Detection of microcalcifications associated with cancer. In Yaffe, M.J. editor, *Digital Mammography 2000*, in Proc. of the 5th International Workshop on Digital Mammography, Madison, Medical Physics Publishing, pp. 149–153.

Bovis, K. Singh, S 2000. *Detection of masses in mammograms using texture features*, in Proc. of the 15th International Conference on Pattern Recognition, vol. 2, 2267–2270.

Bovis, K. Singh, S. 2002. *Classification of mammographic breast density using a combined classifier paradigm*, in Proc. of the 4th International Workshop on Digital Mammography, pp. 177–180.

Boyle, P. and Levin, B. 2008, *World Cancer Report 2008*, International Agency for Research on Cancer, Lyon, France.

Bozek, J. Delac, K. and Grgic, M. 2008, Computer-Aided Detection and Diagnosis of Breast Abnormalities in Digital Mammography, Proceedings of the 50th International Symposium ELMAR-2008, Zadar, Croatia, pp. 45-52.

Brieman, L. 2001. Random Forests, *Machine Learning*, vol. 5, no. 1, pp. 5–32.

Brown, S. Li, R. Brandt, L. Wilson, L. Kossoff, G. Kossoff, M. 1998, Development of a multi-feature CAD system for mammography. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography*, pp. 189–196. Kluwer, Dordrecht.

Brown, J. Bryan, S. Warren, R. 1996, Mammography screening: an incremental cost effectiveness analysis of double versus single reading of mammograms, *BMJ*, vol. 312, no. 7034, pp. 809–812.

Bruynooghe, M. 2001, High resolution granulometric analysis for early detection of small microcalcification clusters in X-ray mammograms. In Yaffe, M.J. editor, *Digital Mammography 2000*, in Proc. of the 5th International Workshop on Digital Mammography, Madison, Medical Physics Publishing, pp. 154–160.

Brzakovic, D. Luo, X.M. Brzakovic, P. 1990, An approach to automated detection of tumors in mammograms, *IEEE Transactions on Medical Imaging*, vol. 9, no. 3, pp. 233–241.

Burges, C.J.C. 1988. A tutorial on support vector machines for pattern recoginition. *Data Mining and Knowledge Discovery,* vol. 2, pp. 121–167.

Burrel, H.C. Sibbering, D.M. Wilson, A.R.M. Pinder, S.E. Evans, A.J. Yeoman, L.J. Elston, C.W. Ellis, I.O. Blamey, R.W. Robertson, J.F.R. 1996, Screening interval breast cancers: mammographic features and prognostic factors, *Radiology*, vol. 199, pp. 811–817.

Buseman, S. Mouchawar, J. Calonge, N. Byers, T. 2003, Mammgraphy screening matters for young women with breast carcinoma, *Cancer*, vol. 97, no. 1, pp. 352-358.

Byng, J.W. Critten, J.P. Yaffe, M.J. 1997, Thickness-equalization processing for mammographic images. *Radiol*ogy, vol. 203, pp. 564–568.

Byvatov, E. Fechner, U. Sadowski, J. Schneider, G. 2003. Comparison of Support Vector Machine and Artificial Neural Network Systems for Drug/Nondrug Classification, *Journal of Chemical Information and Computer Science*, vol. 43, no. 6, pp. 1882–1889.

Cascio, D. Fauci, F. Magro, R. Raso, G Bellotti, R. De Carlo, F. Tangaro, S. De Nunzio, G. Quarta, M. Forni, G. Lauria, A. Fantacci, M.E. Retico, A. Masala, G.L. Oliva, P. Bagnasco, S. Cheran, S.C. and Torres, E.L. 2008, 'Mammogram Segmentation by Contour Searching and Mass Lesions Classification With Neural Network', *IEEE Transactions on Nuclear Science*, vol. 53, no. 5, pp. 2827-2833.

Cardenosa, G. 1996, Mammography: An overview. In Doi, K. Giger, M.L. Nishikawa, R.M. Schmidt, R.A. editors, *Digital Mammography*, pp. 3–10. Elsevier, Amsterdam.

Cernadas, E. Gomez, L. Rodriguez, P.G. Casas, A. Carrion, R.G. Vidal, J.J. 1996, Design of unsharp masking filters in the frequency domain: Parametrization for breast radiographs. In Doi, K. Giger, M.L. Nishikawa, R.M. Schmidt, R.A. editors, *Digital Mammography*, pp. 463–466. Elsevier, Amsterdam.

Cernadas, E. Zwiggelaar, R. Veldkamp, W. Parr, T. Astley, S. Taylor, C. Boggis, C. 1998, Detection of mammographic microcalcifications using a statistical model. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography 1998*, pp. 205–208. Kluwer, Dordrecht.

Chan, H.P. Doi, K. Galhotra, S. Vyborny, C.J. MacMahon, H. Jokich, P.M. 1987, Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography, *Med Phys*, vol. 14, no. 4 pp. 538–48.

Chan, H.P. Doi, K. Vyborny, C.J.1990. Improvements in radiologists' detection of clustered microcalcifications on mammograms: the potential of computeraided diagnosis, *Investigative Radiology*, vol. 25, pp. 1102–1110.

Chan, H. P. Sahiner, B. Lam, K.L. Petrick, N. Helvie, M.A. Goodsitt, M.M. and Adler, D.D. 1998, 'Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces', *Medical Physics*, vol. 25, no. 10, pp. 2007–2019.

Chan, H.P. Sahiner, B. Helvie, M.A. Petrick, N. Roubidoux, M.A. Wilson, T.E. Adler, D.D. Paramagul, C. Newman, J.S. Sanjay-Gopal, S. 1999, Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study, *Radiology*, vol. 212, pp. 817–827, 1999.

Chang, C.-C. and Lin, C.-J. 2010, LIBSVM: A library for support vector machines. [Online]. Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm

Chen, S.-Y. Lin, W.-C. Chen, C.-T. 1991, Split-and-merge image segmentation based on localized feature analysis and statistical tests, *CVGIP: Graphic. Models Image Processing*. vol. 53, no. 5, pp. 457–475.

Cheng, H-.D. Lui, Y.M. Freimanis, R.I. 1998, A novel approach to microcalcification detection using fuzzy logic technique, *IEEE Trans on Medical Imaging*, vol. 17, no. 3, pp. 442–450.

Chen, Y.W. and Lin, C.J. 2006, Combining SVMs with various feature selection strategies. In Guyon, I. Gunn, S. Nikravesh, M. Zadeh, L., editors, *Feature Extraction, Foundations and Applications*, Springer, New York.

Cheng, H.D. Shi, X.J. Min, R. Hu, L.M. Cai, X.P. Du, H.N. 2006, Approaches for automated detection and classification of masses in mammograms, *Pattern Recognition*, vol. 39, no. 4, pp. 646-668.

Cheevasuvit, F. Maitre, H. Vidal-Madjar, D. 1986, A robust method for picture segmentation based on split-and-merge procedure, *Comput. Vis. Graph. Image Process.*. vol. 34, pp. 268–281.

Chitre, Y. Dhawan, A.P. Moskowitz, M. 1994, Classification of mammographic microcalcifications using image structure and cluster features. In Gale, A.G. Astley, S.M. Dance, D.R. Cairns, A.Y. editors, *Digital Mammography*, pp. 31–40. Elsevier, Amsterdam.

Clausi, D.A. 2002. An analysis of co-occurrence texture statistics as a function of grey level quantization, *Canadian Journal of Remote Sensing*, vol. 28, no. 1, pp. 45–62.

Cortes, C. Vapnik, V. 1995. Support vector networks. *Machine Learning*, vol. 20, pp. 273–297.

Costaridou, L. Sakellaropoulos, P.N. Kristalli, M.A. Skiadopoulos, S.G. Karahaliou, A.N. Boniatis, I.S. Panayiotakis, G.S. 2005. *Multiresolution feature analysis for differentiation of breast masses from normal tissue*, in Proc. of the 1st International Conference on Experiments/Process/System/Modelling and Optimization, Greece, Athens.

Chu, K.C., Smart, C.R. Tarone, R.E. 1988. Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan clinical trial, *Journal of the National Cancer Institute*, vol. 14, pp. 1125-1132.

D'Orsi, C.J. Getty, D.J. Swets, J.A. Pickett, R.M. Seltzer, S.E. and McNeil, B.J. 1992, Reading and decision aids for improved accuracy and standardization of mammographic diagnosis, *Radiology*, vol. 184, pp. 619–622.

Dash M. and Liu, H. 1997, Feature selection for classification, *Intelligent Data Analysis,* vol. 1, no. 3, pp. 131–156.

Day, W.H.E. Edelsbrunner, H. 1984. Efficient algorithms for agglomerative hierarchical clustering methods, *Journal of Classification*, vol. 1, pp. 7–24.

Davies, D.H. Dance, D.R. 1990. Automatic computer detection of clustered calcifications in digital mammograms, *Phys. Med. Biol.*, vol. 35, pp. 1111–1118.

De Koning, H.J. Fracheboud, J. Boer, R. Verbeek, A.L. Collette, H.J. Hendriks, J.H.C.L van Ineveld, B.M. de Bruyn A.E., and van der Maas P.J. 1995, Nation-wide breast cancer screening in Netherlands: support for breast cancer mortaility reduction. National evaluation team for breast cancer screening, *International Journal of Cancer*, vol. 60, no. 6, pp. 777-780.

Dean, P.B. 1996, Overview of breast cancer screening. In Doi, K. Giger, M.L. Nishikawa, R.M. Schmidt, R.A. editors, *Digital Mammography*, pp. 19–26. Elsevier, Amsterdam.

Degroeve, S. Tanghe, K. Baets, B.D. Leman, M. Martens, J.-P. 2005. *A Simulated Annealing Optimization of Audio Features for Drum Classification*, in Proc. of the 6th International Conference on Music Information Retrieval, London, pp. 482–487.

Dehghan, F. Abrishami Moghaddam, H. Giti, M. 2008, "Automatic Detection of Clustered Microcalcifications in Digital Mammograms Study on Applying Adaboost with SVM-based Component Classifiers", Proceedings of the 30th Annual International IEEE EMBS Conference, Vancouver, Canada, pp. 4789-4792.

Dengler, J. Behrens, S. Desaga, J.F. 1993, Segmentation of microcalcifications in mammograms, *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp. 634–642.

Dept. of Biomedical Imaging, Faculty of Medicine, University of Malaya (UM), 2010. [Online]. Available at: http://radiology.um.edu.my/

Dept of Biomedical Imaging, University Malaya Medical Centre (UMMC), Kuala Lumpur, 2010. [Online]. More information available at: http://www.ummc.edu.my/index.php?option=com_content&view=article&id=51:medical-imaging-services&catid=39clinical service s&Itemid=55

Diahi, J.G. Giron, A. Brahmi, D. Frouge, C. Fertil, B. 1998, Evaluation of a neural network classifier for detection of microcalcifications and opacities in digital mammograms. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography 1998*, pp. 151–156. Kluwer, Dordrecht.

Diyana, W.M. Zulaikha, K. and Besar, R. 2002, An Intelligent CAD System for Breast Cancer Detection in Digital Mammograms, International Conference on Artificial Intelligence in Engineering & Technology, Kota Kinabalu Sabah, pp. 476-479.

Diyana, W.M. Besar, R. 2003a, Automated Methods in Clustered Microcalcifications Detection Module of a CAD System, *World Scientific Journals: Journal of Mechanics in Medicine and Biology*, vol. 3, no. 3, pp. 30-33.

Diyana, W.M. Larcher, J. and Besar, R. 2003b, A Comparison of Clustered Microcalcifications Automated Detection Methods in Digital Mammograms. IEEE International Conference on Accoustic, Speech and Signal Processing, pp. 385-388.

Doi, K. Giger, M.L. Nishikawa, R.M. Hoffmann, K.R. MacMahon, H. Schmidt, R. Chua, K.G. 1993, Digital radiography: a useful clinical tool for computer-aided diagnosis by quantitative analysis of radiographic images, *Acta Radiologica*, vol. 34, pp. 426–439.

Domínguez, A.R. Nandi, A.K. 2008. Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection, *Computerized Medical Imaging and Graphics*, vol. 32, no. 4, pp. 304–315.

Domínguez, A.R. Nandi, A.K. 2009a. Toward breast cancer diagnosis based on automated segmentation of masses in mammograms, *Pattern Recognition*, vol. 24, no. 6, pp. 1138–1148.

Domínguez, A.R. Nandi, A.K. 2009b. Development of tolerant features for characterization of masses in mammograms, *Computers in Biology and Medicine*, vol. 39, no. 8, pp. 678–688.

Dror, G. Sorek, R. Shamir, S. 2005. Accurate Identification of Alternatively Spliced Exons using Support Vector Machine, *Bioinformatics*, vol. 21, no. 7, pp. 897–901.

Dua, S. Singh, H. Thompson, H.W 2009. Associative classification of mammograms using weighted rules, *Expert Systems with Applications*, vol. 36, no. 5, pp. 9250–9259.

Duda, R.O. Hart, R.E. Stork, D.G. 2001, *Pattern Classification*. John Wiley & Sons.

Dumais, S. 1998. Using SVMs for Text Categorization, *IEEE Intelligent Systems*, vol. 13, pp. 21–23.

El-Naqa, I. Yang, Y. Wernick, M.N. Galatsanos, N.P. Nishikawa, R.M. 2002, A Support Vector Machine Approach for Detection of Microcalcifications, *IEEE Transactions on Medical Iamging*, vol. 21, no. 12, pp. 1552-1563.

Elmore, J.G. Wells, C.K. Lee, C.H. Howard, D.H. Feinstein. A.R. 1994, Variability in radiologists' interpretations of mammograms, *N Engl J Med*, vol. 331, no. 22, pp. 1493–1499.

Esteve, J. Kricker, A. Ferlay, J. and Parkin, D. 1993, *Facts and figures of cancer in the European Community*, Technical report, International Agency for Research on Cancer, Lyon, France.

Evans, W.P. 1995, "Breast Masses Appropriate Evaluation", *The Radiologic Clinics of North America, Breast Imaging*, vol. 33, no. 6, pp. 1085-1108.

Feature selection tool for LIBSVM written in python language, 2010. [Online]. Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/

Feig, S.A. Yaffe, M.J. 1995, Digital Mammography, Computer-Aided Diagnosis and Telemammography, *The Radiologic Clinics of North America, Breast Imaging*, vol. 33, no. 6, pp. 1205-1230.

Ferlay, J. Autier, P. Boniol, M. Heanue, M. Colombet, M., Boyle, P. 2007, Estimates of the cancer incidence and mortality in Europe in 2006, *Annals of Oncology*, vol. 18, no. 5, pp 581-592.

Ferrari, R.J. Rangayyan, R.M. Desautels, J.E.L. Borges, R.A. and Frere, A.F. 2004, Automatic identification of the pectoral muscle in mammograms, *IEEE Transactions on Medical Imaging*, vol. 23, no. 2, pp. 232–245.

Ferrari, R.J. Rangayyan, R.M. Desautels, J.E.L. Borges, R.A. and Frere, A.F. 2001, Analysis of asymmetry in mammograms via directional filtering with Gabor wavelets, *IEEE Transactions on Medical Imaging*, vol. 20, no. 9, pp. 953–964.

Fisher, R.A. 1936, The use of multiple measurements in taxonomic problems, *Ann. Eugenics*, vol. 7, pp. 178-188.

Fogel, D.B. Wasson, E.C. Boughton, E.M. Porto, V.W. Angeline, P.J. 1998, Linear and neural models for classifying breast masses, *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pp. 485-488.

Frieß, T.-T. N. Cristianini, N. Campbell, C. 1998. *The Kernel-Adatron Algorithm: A Fast and Simple Learning Procedure for Support Vector Machines*, in Proc. of the 15th International Conference on Machine Learning, San Francisco, California, pp. 188–196.

Friedman, P.J. 1999, The past and future of radiological error. In Krupinski, E.A. editor, *Medical imaging 1999: Image perception and performance*, vol. 3663, pp. 2–7.

Friedrich, M. Sickles, E.A. (eds.) 2000. *Radiological diagnosis of breast diseases*. Springer.

Frisell, J. Lidbrink, E. Hellström, L. Rutqvist, L.E. 1997. Followup after 11 years–update of mortality results in the Stockholm mammographic screeningtrial, *Breast Cancer Res Treat*, vol. 45, no. 3, pp. 263-270.

Fukunaga, K. 1990, *Introduction to Statistical Pattern Recognition*. Academic Press.

Fukuoka, D. Kasai, S. Fujita, H. Hara, T. Kato, M. Endo, T. Yoshimura, H. 1998, Automated detection of clustered microcalcifications on digitized mammograms. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography 1998*, pp. 197–200. Kluwer, Dordrecht.

Furey, T.S. Christiani, N. Duffy, N. Bednarski, D.W. Schummer, M. Hauessler, D. 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data, *Bioinformatics*, vol. 16, pp. 906–914.

Glatt, L.A. Longbotham, H.G. Arnow, T.L. Shelton, D. Ravdin. P. 1992, Application of weighted-majority minimum-range filters in the detection and sizing of tumors in mammograms. In Loew, M.H. editors, *Medical Imaging VI: Image Processing*, pp. 477–488.

Giger, L. Yin, F. Doi, K. 1990. Investigation of methods for the computerised detection and analysis of mammographic masses, *SPIE Medical Imaging and Image Processing IV*, vol. 1233, pp. 183–184.

Giger, M.L. Vyborny, C.J. Schmidt , R.A. 1994, Computerized characterization of mammographic masses: analysis of spiculation, *Cancer Letters*, vol. 77, pp. 201–211.

Giger, M. MacMahon, H. 1996, Image processing and computer-aided diagnosis, *Radiologic Clinics of North America*, vol. 24, no. 3, pp. 565-596.

Giger, M.L. 1999, Overview of computer-aided diagnosis in breast imaging. In: *Computer Aided Diagnosis in Medical Imaging*, (Doi K, MacMahon H, Giger ML, Hoffmann KR, eds). (Elsevier, Amsterdam), pp. 167-176.

Goergen, S.K. Evans, J.E. Cohen, G.P.B. Macmillan, J.H. 1997. Characteristics of breast carcinomas missed by screening radiologists, *Radiology*, vol. 204, pp. 131-135.

Golub, T.R. Slonim, D.K. Tamayo, P. Huard, C. Gaansenbeek, M. Mesirov, J.P. Coller, H. Loh, M.L. Downing, J.R. Caliguiri, M.A. Bloomfield, C.D. Lander, E.S. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, *Science*, vol. 286, pp. 531–557.

Gonzalez, R.C. and Woods, R.E. 2002, *Digital Image Processing*. Prentice-Hall, Inc., 2nd edition, ISBN 0-201-18075-8. International Edition.

Gotzsche, P.C. Olsen, O. 2000. Is screening for breast cancer with mammography justifiable?, *Lancet*, vol. 355, pp. 129-134.

Gorgel, P. Sertbas, A. Kilic, N. Ucan, O.N. Osman, O. 2009, Mammographic Mass Classification using Wavelet Based Support Vector Machine, *Istanbul University - Journal of Electrical and Electronics Engineering*, vol. 9, no. 1, pp. 867-875.

Green, D.M. and Swets, J.A. 1966, *Signal detection theory*. Wiley, New York.

Groshong, B.R. Kegelmeyer, W.P. 1996, Evaluation of a hough transform method for circumscribed lesion detection. In Doi, K. Giger, M.L. Nishikawa, R.M. Schmidt, R.A. editors, *Digital Mammography*, pp. 361–366. Elsevier, Amsterdam.

Guillemet, H. Benali, H. Kahn, E. DiPaola, R. 1996, Detection and characterization of microcalcifications in digital mammography. In Doi, K. Giger, M.L. Nishikawa, R.M Schmidt, R.A. editors, *Digital Mammography 1996*, pp 225–230. Elsevier, Amsterdam.

Guliato, D. Rangayyan, R.M. de Carvalho, J.D. Santiago, S. A. 2006, "*Spiculation-Preserving Polygonal Modeling of Contours of Breast Tumors*", in Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), New York, August-September 2006, pp. 2791-2794.

Gulsrud, T.O. Loland, E. 1996. Multichannel filtering for texture extraction in digital mammograms, in Proc. of the *18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Amsterdam, The Netherlands.

Gupta, R. Undrill, P.E. 1995. The use of texture analysis to delineate suspicious masses in mammography, *IOPscience*, vol. 40, pp. 835–855.

Guyon, I. Boser, B. Vapnik, V. 1993. Automatic capacity tuning of very large VC-dimension classifiers, *Advances in Neural Information Processing Systems,* vol. 5 Morgan Kaufmann, San Mateo, CA.

Guyon, I. Eliseeff A. E. 2003, An introduction to variable and feature selection. *Journal of Machine Learning Research*, vol. 3 pp. 1157–1182.

Gürcan, M.N. Yardimci, Y. Cetín, A.E. 1998, Microcalcifications detection using adaptive filtering and Gaussianitiy tests. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography 1998*, pp. 157–164. Kluwer, Dordrecht.

Hadjiiski, L. Sahiner, B. Chan, H.P. Petrick, N. Helvie, M. 1999, Classification of Malignant and Benign Masses Based on Hybrid ART2LDA Approach, *IEEE Transactions on Medical Imaging*, vol. 18, no. 12, pp. 1178-1187

Hagihara, Y. Kobatake, H. Nawano, S. Takeo, H. 2001, Accurate detection of microcalcifications on mammograms by improvement of morphological processing. In Yaffe, M.J. editor, *Digital Mammography 2000*, in Proc. of the 5th International Workshop on Digital Mammography, Madison, Medical Physics Publishing, pp. 193–197.

Hara, T. Yamada, A. Fujita, H. Iwase, T. Endo, T. 2001, Automated classification method of mammographic microcalcifications by using artificial neural network and ACR BI-RADStm Criteria for Microcalcification Distribution. In Yaffe, M.J. editor, *Digital Mammography 2000*, in Proc. of the 5th International Workshop on Digital Mammography, Madison, Medical Physics Publishing, pp. 198–204.

Haralick, R.M. Shanmugam, K. and Dinstein, I., 1973, Textural Features for Image Classification, *IEEE Transactions on Systems, Man and Cybernetics—Part C*, vol. 3, no. 6, pp. 610–621.

Haralick, R.M. 1979. Statistical and structural approaches to texture, *Proceedings of the IEEE*, vol. 67, pp. 786–804.

Hartswood, M. Procter, R. Williams, L.J. 1998, Prompting in practice: How can we ensure radiologists make best use of computer-aided detection systems in screening mammography. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. Van Erning, L.J.T.O. editors, *Digital Mammography*, pp. 363–370. Kluwer, Dordrecht.

Harris, R. Kinsinger, L.S. 2002, Routinely Teaching Breast Self-Examination is Dead. What Does This Mean?, *Journal of the National Cancer Institute*, vol. 94, no. 19, pp. 1420-1421.

Harris, J.R. Henderson, C.I. Hellman, S. Kinne, D.W. 1996, *Breast Diseases*, 2nd edition, JB Lippincott Company, Philadelphia, PA.

Harvey, J.E. Fajardo, L.L. Inis, C.A. 1993, Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blinded interpretation, *AJR*, vol. 161, pp. 1167–1172.

Haykin, S. 1999. *Neural Networks: A Comprehensive Foundation*. Prentice Hall.

Hassanien, A. Slezak, D. 2006. Rough neural intelligent approach for image classification: A case of patients with suspected breast cancer, *International Journal of Hybrid Intelligent Systems*, vol. 3, pp. 205–218.

Hassanien, A. 2007. Fuzzy rough sets hybrid scheme for breast cancer detection, *Image and Vision Computing*, vol. 25, no. 2, pp. 172–183.

Heath, M. Bowyer, K. Kopans, D. Moore, R. Kegelmeyer, W.P. 2001, *The Digital Database for Screening Mammography*, in Proceedings of the Fifth International Workshop on Digital Mammography, Yaffe, M.J. ed., *Medical Physics Publishing*, 2001, pp. 212–218.

Hendee, W.R. Beam, C. and Hendrick, E. 1999, Proposition: all mammograms shouyld be double-read, *Medical Physics*, vol. 26, pp. 115-118.

Highnam, R.P. Brady, J.M. Shepstone, B.J. 1996, Removing the anti-scatter grid in mammography. In Doi, K. Giger, M.L. Nishikawa, R.M Schmidt, R.A. editors, *Digital Mammography 1996*, pp 459–462. Elsevier, Amsterdam.

Highnam, R. and Brady M. 1999, *Mammographic Image Analysis*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Homer, M.J. 1997. *Mammographic interpretation*. The McGraw-Hill Companies, Inc.

Horowitz, S. L. Pavlidis, T. 1974, *Picture segmentation by a directed split-and-merge procedure*, in Proc. of the 2nd International Joint Conference on Pattern Recognition, pp. 424–433.

Howard, D. Roberts, S.C. Ryan, C. Brezulianu, A. 2008. Textural classification of mammographic parenchymal patterns with the sonnet selforganizing neural network, *Journal of Biomedicine and Biotechnology*, vol. 2008, Article ID 526343, 11 pages.

Hsu, C.-N. Huang, H.-J. Dietrich, S. 2002, The ANNIGMA- wrapper approach to fast feature selection for neural nets. *IEEE Trans Systems, Man and Cybernetics,* vol. 32, no. 2, pp. 207–212.

Hsu, C.-W. Chang, C.-C. Lin, C.-J. 2003. *A Practical Guide to Support Vector Classification*. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, 2003.

Huang, G.-B. Babri, H.A. 1998, Upper Bounds on the Number of Hidden Neurons in Feedforward Networks with Arbitrary Bounded Nonlinear Activation Functions, *IEEE Transactions on Neural Networks*, vol. 9, no. 1, pp. 224–229.

Huang, G.-B. 2003, Learning Capability and Storage Capacity of Two-Hidden-Layer Feedforward Networks, *IEEE Transactions on Neural Networks*, vol. 14, no. 2, pp. 274–281.

Huang, S.F. Chang, R.F. Chen, D.R. Moon, W.K. 2004, Characterization of spiculation on ultrasound lesions, *IEEE Transactions on Medical Imaging*, vol. 23, no. 1, pp. 111-121.

Huang, G.-B. Zhu, Q.-Y. Siew, C.-K. 2004, *Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks*, in Proc. of the International Joint Conference on Neural Networks, Budapest, Hungary, vol. 2, pp. 985–990.

Huang, G.-B. and Chee-Kheong, S. 2004, *Extreme Learning Machine: RBF Network Case*, in Proc. of the 8th International Conference on Control, Automation, Robotics and Vision, Kunming, China, vol. 2, pp. 1029–1036.

Huang, G.-B. Zhu, Q.-Y. Siew, C.-K. 2006a, Extreme Learning Machine: Theory and Applications, *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501.

Huang, G.-B. Chen, L. Siew, C.-K. 2006b, Universal Approximation using Incremental Constructive Feedforward Networks with Random Hidden Nodes, *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892.

Huang, G.-B. Zhu, Q.-Y. Mao, K. Z. Siew, C.-K. Saratchandran, P. Sundararajan, N. 2006c, Can Threshold Networks be Trained Directly?, *IEEE Transactions on Circuits and Systems - II: Express Briefs*, vol. 53, no. 3, pp. 187–191.

Huo, Z. Giger, M.L. Vyborny, C.J. Bick, U. Lu, P. Wolverton, D.E. Schmidt, R.A. 1995, Analysis of spiculation in the computerized classification of mammographic masses, *Med Phys*, vol. 22, pp. 1569–1579.

Huo, Z. Giger, M.L. Vyborny, C.J. and Metz, C.E. 2000, Effectiveness of CAD in the diagnosis of breast cancer: an observer study on an independent database of mammograms, *Radiology*, vol. 7, pp. 1077–1084.

Hutt, I. 1996, *The computer-aided detection of abnormalities in digital mammograms*. PhD thesis, Faculty of Medicine, Department of Medical Biophysics, University of Manchester.

IARC Handbooks of Cancer Prevention 2002, *Beast Cancer Screening*, vol. 7, Lyon: IARC Press 2002.

Ibrahim, N. Fujita, H. Hara, T. Endo, T. 1997. Automated detection of clustered microcalcifications on mammograms: CAD system application to MIAS database, *Physics in Medicine and Biology*, vol. 42, no. 12, pp. 2577–2589.

Image Processing Toolbox (R2009b), MATLAB (Matrix Laboratory). Mathworks, 2009. [Online]. Available at: http://www.mathworks.com/help/toolbox/images/

Jackson, V.P. 2002. Screening mammography: controversies and headlines. *Radiology,* vol. 225, no. 2, pp. 323-326.

Jiang, Y. Nishikawa R.M. Wolverton, E.E. Metz, C.E. Giger, M.L. Schmidt, R.A. and Vyborny, C.J. 1996, Malignant and benign clustered microcalcifications: automated feature analysis and classification, *Radiology*, vol. 198, pp. 671–678.

Jiang, Y. Nishikawa, R.M. Schmidt, R.A. Metz, C.E. Giger, M.L. Doi, K. 1998, Benefits of computer-aided diagnosis (CAD) in mammographic diagnosis of malignant and benign clustered microcalcifications. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography 1998*, pp. 215–220. Kluwer, Dordrecht.

Jiang, Y. Nishikawa, R.M. Schmidt, R.A. Metz, C.E. Giger, M.L. and Doi, K. 1999, Improving breast cancer diagnosis with computer-aided diagnosis, *Academic Radiol.*, vol. 6, pp. 22–33.

Jirari, M. 2005, "A Computer Aided Detection System for Digital Mammograms Based on Radial Basis Functions and Feature Extraction Techniques", Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, pp. 4457-4460.

Joachims, T. 1998. *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*, in Proc. of the 10th European Conference on Machine Learning, Springer Verlag, Heidelberg, pp. 137–142.

Juhl. J. 1982. *Paul and Juhl's Essentials of Roentgen Interpretation*. 4th edition, Harper & Row, Philadelphia, pp. 340–345.

Karssemeijer, N. 1992, A stochastic model for automated detection of calcifications in digital mammograms, *Image and Vision Computing*, vol. 10, pp. 369–375.

Karssemeijer, N. 1993, Recognition of clustered microcalcifications using a random field model. In Acharya, R.S. Goldgof, D.B. editors, *Biomedical Image Processing and Biomedical Visualization*, vol. 1905, pp. 776–786.

Karssemeijer, N. and te Brake, G.M. 1996, Detection of stellate distortions in mammograms, *IEEE Transactions on Medical Imaging,* vol. 15, pp. 611–619.

Karssemeijer. N. and te Brake, G. 1998 Combining single view features and asymmetry for detection of mass lesions. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography*, pp. 95–102, Nijmegen: Kulwer, Academic Publishers,

Karssmeijer, N. Otten, J.D. Verbeek, A.L. Groenewoud, J.H. De Koning, H.J. Hendriks, J.H Holland, R. 2003. Computer-aided detection versus independent double reading of masses on mammograms. *Radiology*, vol. 227, no. 1, pp. 192-200.

Karahaliou, A. Boniatis, I. Skiadopoulos, S. Sakellaropoulos, P. Likaki, E. Panayiotakis, G. Costaridou, C. 2006. *A texture analysis approach for characterizing microcalcifications on mammograms*, in Proc. of the *International Special Topic Conference on Information Technology* in Biomedicine, Greece.

Karahaliou, A. Boniatis, I. Sakellaropoulos, P. Skiadopoulos, S. Panayiotakis, G. Costaridou, L. 2007. Can texture of tissue surrounding microcalcifications in mammography be used for breast cancer diagnosis, *Nuclear Instruments and Methods in Physics Research*, vol. 580, no. 2, pp. 1071–1074.

Karahaliou, A. Boniatis, I. Skiadopoulos, G. Sakellaropoulos, F. Arikidis, N. Likaki, E.A. Panayiotakis, G. Costaridou, L. 2008. Breast cancer diagnosis: Analyzing texture of tissue surrounding microcalcifications, *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 6, pp. 731–738.

Kaufman, L. Rousseeuw, P. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. J. Wiley, New York.

Kaufmann, G.H. Salfity, M.F. Granitto, P. Ceccatto, H.A. 2001, Automated detection and classification of clustered microcalcifications using morphological filtering and statistical techniques. In Yaffe, M.J. editor, *Digital Mammography 2000*, in Proc. of the 5th International Workshop on Digital Mammography, Madison, Medical Physics Publishing, pp. 253–258.

Kearns, M. Mansour, Y. Ron, D. 1997. An experimental and theoretical comparison of model selection methods, *Machine Learning*, vol. 27, pp. 7–50.

Kegelmeyer, W.P. Pruneda, J.M. Bourland, P.D. Hillis, A. Riggs, M.W. Nipper, M.L. 1994, Computer-aided mammographic screening for spiculated lesions, *Radiology*, vol. 191, pp. 331–337.

Kegelmeyer Jr., W.P. 1994. Evaluation of stellate lesion detection in a standard mammogram data set, *Int. J. Pattern Recogn. Artificial Intell.*, vol. 7, no. 12, pp. 1477–1493.

Khuzi, A.M. Besar, R. Zaki, W.M.D.W. Ahmad, N.N. 2009, Identification of masses in digital mammogram using gray level co-occurrence matrices, *Biomedical Imaging and Intervention Journal*, vol. 5, no. 3, pp. 1-13.

Kim, J.K.K. Min, B. 2002, Classification of Malignant and Benign Tumors Using Boundary Characteristics in Breast Ultrasonograms, *Journal of Digital Imaging*, vol. 15, pp. 224-227.

Kobatake, H. Takeo, H. Nawano, S. 1998, Microcalcification detection system for full-digital mammography. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography 1998*, pp. 201–204. Kluwer, Dordrecht.

Kocur, C.M. Rogers, S.K. Myers, L.R. Bums, T. Kabrisky, M. Hoffmeister, J.W. Bauer, K.W. Steppe, J.M. 1996, Using neural networks to select wavelet features for breast cancer diagnosis, *IEEE Engineering In Medicine And Biology Magazine*, May/June, pp. 95-105.

Kohavi, R. 1995. *A Study of cross-validation and bootstrap for accuracy estimation and model selection*, in Proc. of the International Joint Conference on Artificial Intelligence, pp. 1137–1145.

Kohavi, R. and John, G. 1997. Wrappers for feature subset selection, *Artificial Intelligence Journal*, vol. 97, pp. 273–324.

Koller, D. and Sahami, M. 1996, *Toward Optimal Feature Selection*, in Proc. of the Thirteenth International Conference on Machine Learning, ICML, pp. 284–292.

Kopans, D.B. 1989, *Breast Imaging.* J. B. Lippincott Company.

Kopans D.B. 1992, The positive predictive value of mammography, *Amer. J. Roentgenol.*, vol. 158, pp. 521–526, 1992.

Kramer, D. and Aghdasi, F. 1999, *Texture analysis techniques for the classification of microcalcifications in digitized mammograms*, in Proc. of the 5th IEEE AFRICON Conference, Cape Town, Africa, pp. 395–400.

Krupinski, E.A. Nodine, C.F. 1994, Gaze duration predicts the locations of missed lesions in mammography. In Gale, A.G. Astley, S.M. Dance, D.R. Cairns, A.Y. editors, *Digital Mammography*, pp. 399–405. Elsevier, Amsterdam.

Kundel, H.L. Nodine, C.F. 1978, Studies of eye movements and visual search in radiology. In Seders, J.A.W. Fisher, D. Monty, R. editors, *Eye movements and the higher psychological functions*. Hillsdale, New Jersey.

Kwok, S. Z. Chandrasekhar, R. Attikiouzel, Y. and Rickard, M.T. 2004, Automatic pectoral muscle segmentation on mediolateral oblique view mammograms, *IEEE Transactions on Medical Imaging*, vol. 23, no. 9, pp. 1129–1140.

Lai, S.M. Li, X. Bischof, W.F. 1989, On techniques for detecting circumscribed masses in mammograms, *IEEE Trans. Medical Imaging*, vol. 8, no. 4, pp. 377–386.

Laming, D. 1995, Screening cervival smears, *British Journal of Psychology*, vol. 86, pp. 507–516.

Lanyi, M. 1988, *Breast calcifications*: Springer-Verlag, Berlin, Heidelberg.

Lasdon, L.S. 1970. *Optimization Theory for Large Systems*, MacMillan.3

Lau, T.K. and Bischof, W.F. 1991, Automated detection of breast tumors using the asymmetry approach, *Comp and Biomed Research*, vol. 24, pp. 273–295.

Laws, K.I. 1980, *Textured Image Segmentation*. PhD thesis, University of Southern California, United States.

LeCun, Y. 1986. Learning processes in an asymmetric threshold network. In *Disordered Systems and Biological Organizations*, pp. 233–240, Springer-Verlag, Les Houches, France.

Lee, S.K. Lo, C.S. Wang, C.M. Chung, P.C. Chang, C.I. Yang, C.W. Hsu, P.C. 2000, A computer-aided design mammography screening system for detection and classification of microcalcifications, *International Journal of Medical Informatics*, vol. 60, pp. 29–57.

Lee, R. Alberdi, E. Taylor, P. 2001, A comparative study of four techniques for calcification detection. In Yaffe, M.J. editor, *Digital Mammography 2000*, in Proc. of the 5th International Workshop on Digital Mammography, Madison, Medical Physics Publishing, pp. 264–271.

Lerner, B.H. 2002, When statistics provide unsatisfying answers: revisiting the breast selfexamination controversy, *Canadian Medical Association Journal*, vol. 166, no. 2, pp. 199-201.

Levi, F. Lucchini, F. Negri, E. Boyle, P. La and Vecchia, C. 2003, 'Mortality from major cancer sites in the European Union', *Annals of Oncology*, vol. 14, pp. 490-495.

Levi, F. Bosetti, C. Lucchini, F. Negri, E. La Vecchia, C. 2005, Monitoring the decrease in breast cancer mortality in Europe, *European Journal of Cancer Prevention*, vol. 14, no. 6, pp. 497–502.

Lewin, J.M. Hendrick, R.E. D'orsi, C.J. Isaacs, P.K. Moss, L.J. Karellas, A. Sisney, G.A. Kuniu, C.C. Cutter, G.R. 2001. Comparison of full-field digital mammography with screen-film mammography for cancer detection: results of 4,945 paired examinations, *Radiology*, vol. 218, no. 3, pp. 873-880.

Li, H. Lui, K.J. Lo, S.C.B. 1997, Fractal modelling and segmentation for the enhancement of microcalcifications in digital mammograms, *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 785–798.

Liang, N.-Y. Huang, G.-B. Saratchandran, P. Sundararajan, N. 2006, A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks, *IEEE Transactions on Neural Networks*, vol. 17, no. 6, pp. 1411–1423.

Lim, J. S. 1990. Two-Dimensional Signal and Image Processing, Englewood Cliffs, NJ, Prentice Hall, pp. 469–476.

Lim, G.C.C. Rampal, S. Halimah, Y. (eds.) 2008, *Cancer Incidence in Peninsular Malaysia 2007-2008*. National Cancer Registry, Kuala Lumpur.

Lloyd, S.P. 1982, Least squares quantization in PCM, *IEEE Transactions on Information Theory*, vol. 2, pp. 129–137.

Lladó, X. Oliver, A. Martì, J. Freixenet, J. 2007. *Dealing with false positive reduction in mammographic mass detection*, in Proc. of the National Conference on Medical Image Understanding and Analysis UK, pp. 81–85, Aberystwyth, Wales, UK.

Lladó, X. Oliver, A. Freixenet, J. Martì, R. Martì, J. 2009. A textural approach for mass false positive reduction in mammography, *Computerized Medical Imaging and Graphics*, vol. 33, no. 6, pp. 415–422.

Lu, S. Bottema, M.J. 2001, Classifying lobular and DCIS microcalcification. In Yaffe, M.J. editor, *Digital Mammography 2000*, in Proc. of the 5th International Workshop on Digital Mammography, Madison, Medical Physics Publishing, pp. 280–284.

Lyra, M. Lyra, S. Kostakis, B. Drosos, S. Georgosopoulos, C. 2008. *Digital mammography texture analysis by computer assisted image processing*, in Proc. of the IEEE International Workshop on Imaging, Chania, Greece, pp. 223–227.

Makinacı, M. 2005, Support Vector Machine Approach for Classification of Cancerous Prostate Regions, *World Academy of Science, Engineering and Technology*, no. 7, August, pp. 166-169.

Manning, D.J. Ethell, S.C. Donovan, T. 2004. Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph. *Br. J. Radiol.*, vol. 77, no. 915, pp. 231-235.

Manzano-Lizcano, J. A. Sánchez-Ávila, C. Moyano-Pérez. L. 2004, "A Microcalcification Detection System for Digital Mammography using the Contourlet Transform", in Proceedings of the 2004 International Conference on Computational & Experimental Engineering & Science, 26-29 July, Madeira, Portugal, pp. 611-616.

Mao, J. Jain, A.K. 1996. A self-organizing network for hyperellipsoidal clustering (HEC). *IEEE Transactions on Neural Networks*, vol. 7, pp. 16–29.

Martins, L.de.O. Junior, G.B. Silva, A.C. Paiva, A.C.de. Gatass, M. 2009, Detection of Masses in Digital Mammograms using K-means and Support Vector Machine, *Electronic Letters on Computer Vision and Image Analysis*, vol. 8, no. 2, pp. 39-50.

Martì, R. Zwiggelaar, R. Rubin, C. 2000. *A novel similarity measure to evaluate image correspondence*, in Proc. of the 15th International Conference on Pattern Recognition, vol. 3, pp. 3171–3174.

Martí, J. Freixenet, J. García, R. Español, J. Golobardes, E. Salamó, M. 2001, Classification of microcalcifications in digital mammograms using case-based reasoning. In Yaffe, M.J. editor, *Digital Mammography 2000*, in Proc. of the 5th International Workshop on Digital Mammography, Madison, Medical Physics Publishing, pp. 285–294.

Martì, J. Freixenet, J. Noz, X.M. Oliver, A. 2003. *Active region segmentation of mammographic masses based on texture, contour and shape features*. Springer-Verlag Berlin Heidelberg, Lecture Notes on Computer Science, vol. 2652, pp. 478–485.

Meesman, D. Scheunders, P. VanDyck, D. 1998, Classification of microcalcifications using texture-based features. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography 1998*, pp. 233–236. Kluwer, Dordrecht.

Mercer, J. 1909. Functions of positive and negative type and their connection with the theory of integral equations, *Transactions of the London Philosophical Society*, vol. 29, pp. 415–446.

Miller, P. Astley, S. 1992. Classification of breast tissue by texture analysis, *Image and Vision Computing*, vol. 10, pp. 277–283.

Miller, A.B. Baines, C.J. To, T. Wall, C. 1992a. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years, *CMAJ*, vol. 147, no. 10, pp. 1459-1476.

Mirzaalian, H. Ahmadzadeh, M.R. and Sadri, S. 2007, *Pectoral Muscle Segmentation on Digital Mammograms by Nonlinear Diffusion Filtering*, in Proc. of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 22-24 Aug., pp. 581–584.

Monsees, B.S. 1995, Evaluation of Breast Microcalcifications, *The Radiologic Clinics of North America, Breast Imaging*, vol. 33, no. 6, pp. 1109-1121.

Morrow, W.M. Paranjape, R.B. Rangayyan, R.M. Desautels, J.E.L., 1992, "Region-based contrast enhancement of mammograms", *IEEE Transactions on Medical Imaging*, vol. 11, no. 3, pp. 392–406.

Moskowitz, M. 1989, Impact of a priory medical detection on screening for breast cancer, *Radiology*, vol. 184, pp. 619–622.

Mousa, R. Munib, Q. Moussa, A. 2005, Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. *Expert Systems With Applications*, vol. 28, no. 4, pp. 713-723.

Mudigonda, N.R. Rangayyan, R. Desautels, J.E.L. 2000, Gradient and Texture Analysis for the Classification of Mammographic Masses, *IEEE Transactions on Medical Imaging*, vol. 19, no. 10, pp. 1032-1043.

Mudigonda, N.R. Rangayyan, R.M. and Desautels, J.E.L. 2001, Detection of Breast Masses in Mammograms by Density Slicing and Texture Flow-Field Analysis, *IEEE Transactions on Medical Imaging,* vol. 20, no. 12, pp. 1215-1227.

Mutihac, R. Colavita, AA. Cicuttin, A. Cerdeira, A. 1998, Maximum entropy improvement of x-ray digital mammograms. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography*, pp. 329–336. Kluwer, Dordrecht.

Müller, K.-R. Mika, S. Ratsch, G. Tsuda, K. Scholkopf, B. 2001, An introduction to kernel-based learning algorithms, *IEEE Trans on Neural Networks*, vol. 12, no. 2, pp. 181–201.

Nagel, R.H. Nishikawa, R.M. Papaioannou, J. Doi, K. 1998, Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms, *Med. Phys.*, vol. 25, no. 8, pp. 1502–1506.

Netsch, T. Biel, M. Peitgen, H.O. 1998, Display of high-resolution digital mammograms on crt monitors. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography*, pp. 313–320. Kluwer, Dordrecht.

Newcomb, P.A. Weiss, N.S. Storer, B.E. Scholes, D. Young, B.E. Voigt, L.F. 1991, Breast self-examination in relation to the occurrence of advanced breast cancer, *J Nat Cancer Inst*, vol. 83, no. 4, pp. 260–265.

Ng, S.L. and Bischof, W.F. 1992, Automated detection and classification of breast tumors, *Comput Biomed Res.*, vol. 25 pp. 218–237

Nicolaou, N. Petroudi, S. Georgiou, J. Polycarpou, M. and Brady, M. 2008, *Digital mammography: Towards pectoral muscle removal via Independent Component Analysis*, in Proc. of 4th IET International Conference on Advances in Medical, Signal and Information Processing, 14-16 Jul., pp. 1–4.

Nishikawa, R.M. Giger, M.L. Doi, K. Vyborny, C.J. Schmidt, R.A. 1994, Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms. In Bowyer, K.W. Astley, S.M. editors, *State of the art in digital mammographic image analysis*, vol. 9 of series in machine perception and artificial intelligence, World Scientific, pp. 82–102.

Novikoff, A.B.J. 1962. On convergence proofs on perceptrons, in Proc. of the Symposium on the Mathematical Theory of Automata, vol. 12, pp. 615–662.

Oliver, A. Freixenet, J. Bosch, A. Raba, D. Zwiggelaar, R. 2005. Automatic classification of breast tissue. Springer-Verlag, Berlin, Heidelberg, pp 431–438.

Oliver, A. Lladó, X. Martì, R. Freixenet, J. Zwiggelaar, R. 2007a. *Classifying mammograms using texture information*, in Proc. of the National Conference on Medical Image Understanding and Analysis UK, pp. 223–227, Aberystwyth, Wales, UK.

Oliver, A. Lladó, X. Martì, R. Freixenet, J. Martì, J. 2007b. *False positive reduction in mammographic mass detection using local binary patterns*, in Proc. of the Int. Conf. Med. Image Comput. Comput. Assist. Interv., vol. 4478, pp. 286–293.

Olsen, O. Gotzsche, P.C. 2001. Cochrane review on screening for breast cancer with mammography. *Lancet*, vol. 358, pp. 1340-1342.

Osareh, A. Mirmehdi, M. Thomas, B.T. Markham, R. 2002, *Comparative Exudate Classification using Support Vector Machines and Neural Networks*, in Proc. of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention, Tokyo, Japan, pp. 413–420.

Oskoei, M.A. Hu, H. 2008. Support Vector Machine-Based Classification Scheme for Myoelectric Control Applied to Upper Limb, *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1956–1965.

Osuna, E. 1998. Applying SVMs to Face Detection, *IEEE Intelligent Systems*, vol. 13, pp. 23–26.

Özekes, S. Osman, O. Çamurcu, A.Y. 2005. Mammographic Mass Detection Using a Mass Template, *Korean Journal of Radiology*, vol. 6, no. 4, pp. 221–228.

Palmer, G.M. Zhu, C. Breslin, T.M. Xu, F. Gilchrist, K.W. and Ramanujam, N. 2003, Comparison of Multiexcitation Fluorescence and Diffuse Reflectance Spectroscopy for the Diagnosis of Breast Cancer, *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 11, pp. 1233-1242.

Papadopoulosa, A. Fotiadisb, D.I. Likasb, A. 2005, Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines, *Artificial Intelligence in Medicine*, vol. 32, no. 2, pp. 141-150.

Parkin, D. M. Bray, F. Ferlay, J. Pisani, P. 2005, Global Cancer Statistics, 2002, *CA: A Cancer Journal for Clinicians*, vol. 55, no. 74.

Parr, T.C. Astley, S.M. Taylor, C.J.Boggis, C.R.M. 1996a, Model based classification of linear structures in digital mammograms.

Parr, T.C. Taylor, C.J. Astley, S.M. Boggis, C.R.M. 1996b, A statistical representation of pattern structure for digital mammography. In Doi, K. Giger, M.L. R M Nishikawa, R.M. Schmidt, R.A. editors, *Digital Mammography*, pp. 357–360. Elsevier, Amsterdam.

Pavlidis, T. Liow, Y.-T. 1990, Integrating region growing and edge detection, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 225–233.

Peer, P.G.M. Werre, J.M. Mravunac, M. Hendriks, J.H.C.L. Holland, R. Verbeek, A.L.M. 1995, Effect on breast cancer mortality of biennial mammographic screening of women under age 50, *Int. J. Cancer*, vol. 60, pp. 808–811.

Pfisterer, R. Aghdasi, R. 1998. Detection of masses in digitised mammograms, in Proc. of the 1998 South African Symposium on Communications and Signal Processing, South Africa, pp. 115–120.

Pfisterer, R. Aghdasi, F. 1999, *Comparison of texture based algorithms for the detection of masses in digitized mammograms*, in Proc. of the IEEE AFRICON, Cape Town, South Africa, vol. 1, pp. 383–388.

Pfisterer, R. Aghdasi, F. 2001. Tumor detection in digitized mammograms by image texture analysis. *Society of Photo-Optical Instrumentation Engineers*, vol. 40, pp. 209–216.

Pisano, E.G. Gatsonis, C. Hendrick, E. Yaffe, M. Baum, J.K. Acharyya, S. Conant, E.F. Fajardo, L.L. Basett, L. D'orsi, C.J. Jong, R. Rebner, M. 2005. Diagnostic performance of digital versus film mammography for breast-cancer screening, *N Engl J Med*, vol. 353, no. 17, pp. 1773-1783.

Platt, J.C. 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, Technical Report MSR-TR-98-14, Microsoft Research Center, Redmond, United States.

Platt, J.C. 1999a. Fast Training of Support Vector Machines using Sequential Minimal Optimization. In Smola, A.J. Bartlett, P.L. Schölkopf, B. Schuurmans, D. editors, *Advances in Large Margin Classifiers*, pp. 185–208, MIT Press, Cambridge, Massacheusets.

Platt, J.C. 1999b. Probabilistic Output for Support Vector Machines and Comparison to Regularized Likelihood Methods. In Smola, A.J. Bartlett, P.L. Schölkopf, B. Schuurmans, D. editors, *Advances in Large Margin Classifiers*, pp. 61–74, MIT Press, Cambridge, Massacheusets.

Pohlman, S. Powell, K.A. Obuchowski, N.A. Chilcote, W.A. Grundfest-Broniatowski, S. 1996, Quantitative classification of breast tumors in digitized mammograms, *Med Phys.*, vol. 23, pp. 1337–1345.

Polakowski, W.E. Cournoyer, D.A. Rogers, S.K. DeSimio, M.P. Ruck, D.W. Hoffmeister, J.W. Raines, R.A. 1997, Computer-aided breast cancer detection and diagnosis of masses using difference of gaussians and derivative-based feature saliency, *IEEE Trans on Medical Imaging*, vol. 16, no. 6, pp. 811–819.

Popli, M.B. 2001, Pictorial essay: Mammographic features of breast cancer, *Ind J. Radiol. Imag.,* vol. 11, pp. 175–179.

Qian, W.C.L. Kallergi. M. Clark, R.A. 1994, Tree-structured nonlinear filters in digital mammography, *IEEE Trans. Medical Imaging, vol. 13*, no.1, pp. 25–36.

Raba, D. Oliver, A. Martí, J. Peracaula1, M. and Espunya, J. 2005, *Breast Segmentation with Pectoral Muscle Suppression on Digital Mammograms*, in Lecture Notes on Computer Science, vol. 3523, pp. 471–478.

Rahbar, G. Sie, A.C. Hansen, G.C. Prince, J.S. Melany, M.L. Reynolds, H.E. Jackson, V.P. Sayre, J.W. Bassett, L.W. 1999. Benign versus malignant solid breast masses: US differentiation, *Radiology*, vol. 213, no. 3, pp. 889-894.

Rangayyan, R.M. El-Faramawy, N.M. Desautels, J.E.L. Alim, O.A. 1997, Measures of Acutance and Shape for Classification of Breast Tumors, *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 799-810.

Rangayyan, R.M. Desautels, J.E.L. 2000, Boundary modelling and shape analysis methods for classification of mammographic masses, *Medical and Biological Engineering and Computing*, vol. 38, no. 5, pp. 487-496.

Rangayyan, R.M. 2005. *Biomedical Image Analysis*. Biomedical Engineering Series. CRC Press LLC, ISBN 0-8493-9695-6.

Richards, W. Polit, A. 1974. Texture matching. *Kybernetic*, vol. 16, pp. 155–162.

Ripley, B.D. 1996, *Pattern Recognition and Neural Networks*. Cambridge University Press.

Roebuck, E.J. and Blamey, R.W. 1990, *Clinical Radiology of the Breast*. London: Heinemann Medical Books.

Rosenblatt, F. 1962. *Principles of Neurodynamics*. Spartan Books, Washintgon DC.

Rumelhart, D.E. Hinton, G.E. Williams, R.J. 1986. Learning internal representations by error propagaton. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* vol. 1, pp. 318–362.

Rumelhart, D.E. Hinton, G.E. Williams, R.J. 1986, Learning representations by back-propagating errors, *Nature*, vol. 323, pp. 533–536.

Sahiner, B. Chan, H.P. Petrick, N. Wei, D. Helvie, M.A. Adler, D.D. Goodsitt, M.M. 1996, Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images, *IEEE Trans Med Imag*., vol. 15, pp. 598–610.

Sallam, M. and Bowyer, K.W. 1994, Registering time-sequences of mammograms using a two-dimensional unwarping technique. In Gale, A.G. Astley, S.M. Dance, D.R. Cairns, A.Y. editors, *Digital Mammography*, pp. 121–131. Elsevier, Amsterdam.

Sample, J.T. 2005, *Computer Assisted Screening Of Digital Mammogram Images*. PhD thesis, Dept. of Computer Science, Louisiana State University, USA.

Savage, C.J. Gale, A.G. Pawley, E.F. Wilson, A.R.M. 1994, To err is human; to compute divine? In Gale, A.G. Astley, S.M. Dance, D.R. Cairns, A.Y. editors, *Digital Mammography*, pp. 405–414. Elsevier, Amsterdam.

Schmidt, F. Hartwagner, K.A. Spork, E.B. Groell, R. 1998, Medical audit after 26, 711 breast imaging studies: improved rate of detection of small breast carcinomas, *Cancer*, vol 83, no. 12, pp. 2516–2520.

Schmidt, R.A. Nishikawa, R.M. Osnis, R.B. Schreibman, K.L. Giger, M.L. Doi, K. 1996, Computerized detection of lesions missed by mammography. In Doi, K. Giger, M.L. Nishikawa, R.M. Schmidt, R.A. editors, *Digital Mammography*, pp. 105–110. Elsevier, Amsterdam.

Schölkopf, B. 1997. *Support Vector Learning*. PhD thesis, Technischen Universität Berlin. Published by: R. Oldenbourg Verlag, Munich.

Selvan, S.E. Xavier, C.C. Karssemeijer, N. Sequiera, J. Cherian, R.A. Dhala, B.Y. 2006. Parameter Estimation in Stochastic Mammogram Model by Heuristic Optimization Techniques, *IEEE Trans. on Information Technology in Biomedicine,* vol. 10, no. 4, pp. 685–695.

Shapiro, S. Strax, P. Venet, L. 1971, Periodic breast cancer screening in reducing mortality from breast cancer, *JAMA*, vol. 215, no. 11, pp. 1777–1785.

Shapiro, S. Venet, W. Strax, P. Venet, L. Roeser, R. 1982, Ten- to fourteen-year effect of screening on breast cancer mortality, *Journal of the National Cancer Institute, vol. 69*, pp. *349*-55.

Sheshadri, H.S. Kandaswamya, A. 2007. Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms, *Computerized Medical Imaging and Graphics*, vol. 31, no. 1, pp. 46–48.

Sickles, E.A. 1984, Mammographic features of early breast cancer, *American Journal of Roentgenology*, vol. 143, pp. 461-464.

Sickles, E.A. 1986, Mammographic features of 300 consecutive nonpalpable breast cancers, *American Journal of Roentgenology*, vol. 146, pp. 661-663.

Sickes, E.A. 1997, Breast cancer screening outcomes in women ages 40-49: clinical experience with service screening using modern mammography, *Journal of the National Cancer Institute: Monographs*, vol. 22, pp. 99-104.

Smith, R.A. Caleffi, M. Albert, U.S. Chen, T.H.H. Duffy, S.W. Franceschi, D. Nystrom, L. 2006, Breast Cancer in Limited-Resource Countries: Early Detection and Access to Care, *Breast Journal* vol. 12, pp. 16-26.

Smola, A.J. Schölkopf, B. 2004. A Tutorial on Support Vector Regression, *Statistics and Computing*, vol. 14, no. 3, 2004, pp. 199–222.

Soh, L.-K. Tsatsoulis, C. 1999. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 780–795.

Soltanian-Zadeh, H. Rafiee-Rad, F. and Pourabdollah-Nejad, S. 2004, 'Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms', *Pattern Recognition*, vol. 37, no. 10, pp. 1973–1986.

Song, E. Jiang, L. Jin, R. Zhang, L. Yuan, Y. Li, Q. 2009. Breast Mass Segmentation in Mammography Using Plane Fitting and Dynamic Programming, *Academic Radiology*, vol. 16, no. 7, pp. 826–835.

Sorantin, E. Schmidt, F. Mayer, H. Winkler, P. Szepesvari, C. Graif, E. Schuetz, E. 1998, Automated detection and classification of microcalcifications in mammograms using artificial neural nets. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography*, pp. 225–232. Kluwer, Dordrecht.

Statistics Toolbox (R2009b), MATLAB (Matrix Laboratory). Mathworks, 2009. [Online]. Available at: http://www.mathworks.com/help/toolbox/stats/

Strickland, R.N. 1996a, Wavelet transform for detecting microcalcifications in mammograms, *IEEE Transactions on Medical Imaging*, vol. 15, no. 2, pp. 218–229.

Strickland, R.N. Baig, L.J. Dallas, W.J. Krupinski, E.A. 1996, Wavelet-based image enhancement as an instrument for viewing cad data. In Doi, K. Giger, M.L. Nishikawa, R.M. Schmidt, R.A. editors, *Digital Mammography*, pp. 441–446. Elsevier, Amsterdam.

Stoutjesdijk, M.J. Boetes, C. Jager, G.J. Beex, L. Bult, P. Hendriks, J.H. Laheij, R.J. Massuger, L. Van Die, L.E. Wobbes, T. Barentsz, J.O. 2001. Magnetic resonance imaging and mammography in women with a hereditary risk of breast cancer, *J Natl Cancer Inst*, vol. 93, no. 14, pp. 1095–1102.

Subashini, T.S. Ramalingama, V. Palanivela, S. 2010. Automated assessment of breast tissue density in digital mammograms, *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 33–43.

Suckling, J. Parker, J. Dance, D. *et al*., 1994. The mammographic image analysis society digital mammogram database. Exerpta Medica, *International Congress Series*, vol. 1069, pp. 375–378. [Online] Available at: http://peipa.essex.ac.uk/info/mias.html

Suzuki, K. Li, F. Sone, S. Doi, K. 2005, Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network, *IEEE Trans on Medical Imaging*, vol. 24, no. 9, pp. 1138–1150.

Svetnik, V. Liaw, A. Tong, C. Wang, T. 2004. Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. In Roli, F. Kittler, J. Windeatt, T., editors, Proceedings of the 5th International Workshopon Multiple Classifier Systems, Lecture Notes in Computer Science vol. 3077, pp. 334–343, Springer.

Tabár, L. Fagerbert C.J.G, Gad, A. Baldetorp, L. Holmberg, L.H. Grontoft, O. Ljungquist, U. Lundstrom, B. Manson, J.C. Erklung, G. Day, N.E. and Pettersson F. 1985, Reduction in mortality from breast cancer after mass screening with mammography. Randomized trial from the breast cancer screening work group of the swedish national board of health and welfare, *Lancet*, vol. 1, pp. 829–832.

Tabár, L. Fagerberg, G. Day, N.E. Holmberg, L 1987, What is the optimum interval between mammographic screening examinations? an analysis based on the latest results of the swedish two-county breast cancer screening trial, *Br J Cancer*, vol. 55, no. 5, pp. 547–551.

Tabár, L. Fagerbert C.J.G, Chen, H.H. Duffy, S.W. Smart, C.R. Gad, A. Smith, R.A. 1995. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. *Cancer* vol. 75, no. 10, pp. 2507–2517.

Tarassenko, L. 1998, *Guide to Neural Computing Applications*, Butterworth-Heinemann.

Thangavel, K. Karnan, M. Kumar, R.S. Mohideen, A.K. 2005a, Automatic Detection of Microcalcification in Mammograms-A Review. *International Journal on Graphics Vision and Image Processing*, vol. 5, no. 5 pp. 31–61.

Thangavel, K. and Karnan, M. 2005b, CAD System for Preprocessing and Enhancement Of Digital Mammograms, *ICGST International Journal on Graphics, Vision and Image Processing (GVIP)*, vol. 5, no. 9, pp. 69–74.

Thomas, D.B. Gao, D.L. Ray, R.M. Wang, W.W. Allison, C.J. Chen, F.L. Porter, P. Hu, Y.W. Zhao, G.L. Pan, L.D. 2002, Randomized Trial of Breast Self-Examination in Shanghai: Final Results, *Journal of the National Cancer Institute*, vol. 94, no. 19, pp. 1445–1457.

Thurfjell, E.L. Lernevall, K.A. and Taube A.A.S 1994, Benefit of independent double reading in a population-based mammography screening program, *Radiology*, vol. 191, pp. 241–244.

Thurfjell, E.L. Lindgren, J.A.L. 1996, Breast cancer survival rates with mammographic screening: similar favorable survival rates for women younger and those older than 50 years, *Radiology*, vol. 201, no. 2, pp. 421–426.

Timp, S. Karssemeijer, N. 2004. A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography, *Med. Phys.*, vol. 31, no. 5, pp. 958–971.

Tyczynski, J.E. Plesko, I. Aareleid, T. Primic-Zakelj, M. Dalmas, M. Kurtinaitis, J. Stengrevics, A. Parkin, D.M. 2004, Breast cancer mortality patterns and time trends in 10 new EU member states: Mortality declining in young women, but still increasing in the elderly, *International Journal of Cancer*, vol. 112, no. 6, pp. 1056–1064.

Underwood, J.C.E. (ed.) 1992. *General and systematic pathology*, chapter 16. Churchill Livingstone.

University of California Irvine (UCI) Machine Learning Repository, Center for Machine Learning and Intelligent Systems, University of California (2010). [Online]. Datasets available at: http://archive.ics.uci.edu/ml/

Vaino, H. Bianchini, F. (eds.) 2002. *IACR handbooks of cancer prevention*. Lyon: IARCPress.

Van Dijck, J.A.M. Verbeek, L.M. Hendriks, J.H.C.L. Holland, R. 1993, The current detectability of breast cancer in a mammographic screening program, *Cancer*, vol. 72, pp. 1933–1938.

Van Dijck, J.A. Verbeek, A.L. Beex, L.V. Hendriks, J.H.C.L. Holland, R. Mravunac, M. Straatman, H. Werre, J.M. 1997, Breast-cancer mortality in a non-randomized trial on mammographic screening in women over age 65, *Int J Cancer*, vol. 70, no. 2, pp. 164–168.

Vapnik, V. Chervonenkis, A.J. 1968. On the uniform convergence of relative frequencies of events to their probabilities, *Dokaldy Akademii Nauk USSR*.

Vapnik, V. Chervonenkis, A.J. 1974. *Theory of Pattern Recognition* (in Russian). Nauka, Moscow, (German Translation: W. Wapnik & A. Tscherwonenkis, *Theorie der Zeichenerkennung*, Akademie–Verlag, Berlin, 1979.

Vapnik, V. 1979. *Estimation of Dependences Based on Empirical Data* (in Russian). Nauka, Moscow, (English translation: Springer Verlag, New York, 1982).

Vapnik, V. 1995. *The Nature of Statistical Learning Theory*: Springer Verlag.

Vapnik, V. 1998. *Statistical Learning Theory*: John Wiley.

Varela, C. Karssemeijer, N. Tahoces, P. 2001. Classification of breast tumors on digital mammograms using laws texture features. Springer-Verlag, Berlin, Heidelberg, *Lecture Notes on Computer Science*, vol. 2208, pp. 1391–1392.

Varela, C. Timp, S. Karssemeijer, N. 2006, Use of border information in the classification of mammographic masses, *Phys. Med. Biol*., vol. 51, no.2, pp. 425-441.

Veldkamp, W.J.H. Karssemeijer, N. 1998, Improved correction for signal dependent noise applied to automatic detection of microcalcifications. In Karssemeijer, N. Thijssen, M.A.O. Hendriks, J.H.C.L. van Erning, L.J.T.O. editors, *Digital Mammography 1998*, pp. 169–176. Kluwer, Dordrecht.

Veldkamp, W.J.H. Karssemeijer, N. Otten, J.D.M. Hendriks, J.H.C.L. 2000, Automated Classification of Clustered Microcalcifications into Malignant and Benign Types, *Medical Physics*, vol. 27, no. 11, pp. 2600–2608.

Verma, B. and Zakos, J. 2001, A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques, *IEEE Transactions on Information Technology in Biomedicine*, vol. 5, no. 1, pp. 46-54.

Veropoulos, K. 2001, *Machine Learning Approaches to Medical Decision Making.* PhD thesis, University of Bristol, United Kingdom.

Vitak, B. 1998, Invasive interval cancers in the Östergötland mammographic screening programme: Radiological analysis, *European Radiology*, vol. 8, pp. 639–646.

Wallet, B.C. Solka, J.L. Priebe, C.E. 1997, A method for detecting microcalcifications in digital mammograms, *Journal of Digital Imaging*, vol. 10, pp. 136–139.

Wang T.C. and Karayiannis, N.B. 1998, Detection of Microcalcifications in Digital Mammograms Using Wavelets*, IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 500–509.

Wang, H.-Q. Huang, D.-S. Wang, B. 2005. Optimisation of Radial Basis Function Classifiers using Simulated Annealing Algorithm for Cancer Classification, *Electronics Letters*, vol. 41, no. 11, pp. 630–632.

Wei, D. Chan, H.P. Helvie, M.A. Sahiner, B. Petrick, N. Adler, D.D. Goodsit, M.M. 1995, Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis, *Med Phys*, vol. 22, pp. 1501–1513.

Wei, L. Yang, Y. Nishikawa, R.M. Jiang, Y. 2005, A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications, *IEEE Transactions on Medical Imaging*, vol. 24, no.5 pp. 371-380.

Wen, J. Zhao, J.L. Luo, S.W. Han, Z. 2000, *The Improvements of BP Neural Network Learning Algorithm*, in Proc. of the 5th International Conference on Signal Processing Proceedings, Beijing, China, vol. 3, pp. 1647–1649.

Wirth, M.A. Lyon, J. Nikitenko, D. Stapinski, A. 2004, Removing Radiopaque Artifacts from Mammograms using Area Morphology, in Proc. of the of SPIE Medical Imaging: Image Processing, pp. 1054–1065.

Wirth, M. Nikitenko, D. Lyon, J. 2007, Segmentation of the Breast Region in Mammograms using a Rule-Based Fuzzy Reasoning Algorithm, *ICGST International Journal on Graphics, Vision and Image Processing (GVIP),* vol. 9, no. 5, pp. 13–22.

Wolberg, W.H. Mangasarian, O.L. 1990, *Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology*, in Proc. of the National Academy of Sciences, U.S.A., vol. 87, Dec. 1990, pp. 9193–9196.

Woods, K. Bowyer, K. 1996, *A general view of detection algorithms*. In Doi, K. Giger, M.L. Nishikawa, R.M. Schmidt, R.A. editors, *Digital Mammography*, pp. 385–390, Elsevier, Amsterdam.

Wu, Y. Giger, M.L. Doi, K. Vyborny, C.J. Schmidt, R.A. and Metz, C.E. 1993, Application of neural networks in mammography: applications in decision making in the diagnosis of breast cancer, *Radiology*, vol. 187, pp. 81–87.

Wu, T.F. Lin, C.-J. Weng, R.C. 2004. Probability Estimates for Multiclass Classification by Pairwise Coupling, *Journal of Machine Learning Research*, vol. 5, pp. 975–1005.

Xu, W. Li, L. and Liu, W. 2007, *A Novel Pectoral Muscle Segmentation Algorithm Based on Polyline Fitting and Elastic Thread Approaching*, in Proc. of the 1st International Conference on Bioinformatics and Biomedical Engineering, 6-8 Jul., pp. 837–840.

Yapa, R.D. Harada, K. 2008, *Breast Skin-Line Estimation and Breast Segmentation in Mammograms using Fast-Marching Method, International Journal of Biomedical Sciences*, vol. 3, no. 1, pp 54–62.

Yin, F.F. Giger, M.L. Doi, K. Metz, C.E. Vyborny, C.J. Schmidt, R.A. 1991, Computerized detection of masses in digital mammograms: Analysis of bilateral substraction images, *Med Phys*, vol. 18 pp. 955–963.

Yin, F.F. Giger, M.L. Vyborny, C.J. Doi, K. Schmidt, R.A. 1993, Comparison of bilateralsubstraction and single-image processing techniques in the computerized detection of mammographic masses. *Invest Radiol*, vol. 6, pp. 473–481.

Yoshida, H. Doi, K. Nishikawa, R.M. 1994, Automated detection of clustered microcalcifications in digital mammograms using wavelet transform techniques, in Proc. of SPIE 2167: Medical Imaging, pp. 868–886.

Yoshida, H. Doi, K. Nishikawa, R. Giger, M. and Schmidt, R. 1996, An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms, *Acad. Radioogy*, vol. 3, no. 8, pp. 621–627.

Youssry, N. Abou-Chadi, F.E.Z. El-Sayad, A.M. 2003. Early detection of masses in digitized mammograms using texture features and neuro-fuzzy model, in Proc. of the 4th Annual IEEE Conf on Information Technology Applications in Biomedicine, Birmingham, United Kingdom, pp. 226–233.

Zhang, M. and Giger, M.L. 1995, Automated detection of spiculated lesions and architectural distortions in digitized mammograms, *SPIE 2434*, pp. 846–855.

Zhao, D. Shridhar, M. Daut, D.G. 1992, *Morphology on detection of calcifications in mammograms*, in Proc. the IEEE International Conference on Acoustics, Speech, and Signal Processing, San Francisco, vol. 3, pp. 129–132.

Zheng, B. Chang, Y.H. Gur, D. 1995, Computerized detection of masses in digitized mammograms using single image segmentation and a multi-layer topographic feature analysis, *Acad Radiol.*, vol. 2, pp. 959–966.

Zonderland, H.M. Coerkamp, E.G. Hermans, J. Van De Vijver, M.J. Van Voorthuisen, A.E. 1999. Diagnosis of breast cancer: contribution of US as an adjunct to mammography, *Radiology*, vol. 213, no. 2, pp. 413-422.

Zucker, S.W. Kant, K. 1981. *Multiple-level representations for texture discrimination*, i*n* Proc of the IEEE Conference on Pattern Recognition and Image Processing, pp. 609–614.

Zwiggelaar, R. Parr, T.C. Schumm, J.E. Hutt, I.W. Taylor, C.J. Astley, S.M. Boggis, C.R.M. 1999, Model-based detection of spiculated lesions in mammography. *Medical Image Analysis*, vol. 3, no. 1, pp. 39–62.

APPENDICES

APPENDIX A

Data Modeling and Analysis

train_features <300x1057 double>

	1	2	3	4	5	6	7	8
1	59.9607	59.9148	60.3982	59.7389	59.9607	59.9148	60.3982	59.7389
2	57.1949	57.0348	57.4103	56.6956	57.1949	57.0348	57.4103	56.6956
3	62.2078	62.1303	62.2055	62.1252	62.2078	62.1303	62.2055	62.1252
4	61.9945	61.9165	62.0680	61.9165	61.9945	61.9165	62.0680	61.9165
5	61.7260	61.6396	61.8298	61.6438	61.7260	61.6396	61.8298	61.6438
6	56.8166	56.5754	56.9408	56.4949	56.8166	56.5754	56.9408	56.4949
7	59.3111	59.0615	59.5478	59.2413	59.3111	59.0615	59.5478	59.2413
8	57.1311	56.7894	57.1276	56.8236	57.1311	56.7894	57.1276	56.8236
9	58.4486	58.2053	58.2913	57.9316	58.4486	58.2053	58.2913	57.9316
10	61.0215	60.8732	61.0742	60.8184	61.0215	60.8732	61.0742	60.8184
11	58.9332	58.6640	58.7441	58.3803	58.9332	58.6640	58.7441	58.3803
12	60.6615	60.4597	60.6181	60.4893	60.6615	60.4597	60.6181	60.4893
13	63.2604	63.1926	63.2389	63.1903	63.2604	63.1926	63.2389	63.1903
14	62.2932	62.1754	62.2775	62.1790	62.2932	62.1754	62.2775	62.1790
15	59.9803	59.8940	60.1870	59.8825	59.9803	59.8940	60.1870	59.8825
16	61.8297	61.7727	61.9825	61.7749	61.8297	61.7727	61.9825	61.7749
17	61.0320	60.9618	61.2471	60.9616	61.0320	60.9618	61.2471	60.9616
18	60.0092	59.9108	60.2167	59.9205	60.0092	59.9108	60.2167	59.9205
19	61.8145	61.7593	62.0029	61.7603	61.8145	61.7593	62.0029	61.7603
20	61.0488	60.9771	61.2633	60.9788	61.0488	60.9771	61.2633	60.9788
21	62.9820	62.8348	62.8678	62.8299	62.9820	62.8348	62.8678	62.8299
22	63.6316	63.5783	63.6025	63.5761	63.6316	63.5783	63.6025	63.5761
23	63.2551	63.1155	63.1498	63.1155	63.2551	63.1155	63.1498	63.1155
24	62.2199	61.8675	61.9186	61.8665	62.2199	61.8675	61.9186	61.8665
25	63.3047	63.0443	63.0658	63.0431	63.3047	63.0443	63.0658	63.0431
26	62.4543	62.1477	62.1951	62.1485	62.4543	62.1477	62.1951	62.1485
27	61.5426	61.1808	61.2569	61.1750	61.5426	61.1808	61.2569	61.1750
28	62.9914	62.6208	62.6524	62.6208	62.9914	62.6208	62.6524	62.6208
29	60.8291	60.3301	60.4456	60.3325	60.8291	60.3301	60.4456	60.3325
30	61.6348	61.4482	61.5813	61.4456	61.6348	61.4482	61.5813	61.4456
31	62.8234	62.6475	62.7556	62.6479	62.8234	62.6475	62.7556	62.6479
32	61.4684	61.2263	61.3831	61.2221	61.4684	61.2263	61.3831	61.2221
33	60.5795	60.5158	60.8625	60.5172	60.5795	60.5158	60.8625	60.5172
34	62.2633	62.2265	62.4956	62.2288	62.2633	62.2265	62.4956	62.2288
35	61.0641	60.9992	61.3215	60.9992	61.0641	60.9992	61.3215	60.9992
36	60.9063	60.6887	61.0800	60.9871	60.9063	60.6887	61.0800	60.9871
37	63.1006	63.0411	63.2598	63.1434	63.1006	63.0411	63.2598	63.1434
38	61.9738	61.8226	62.1431	62.0164	61.9738	61.8226	62.1431	62.0164
39	59.8443	59.6256	59.9599	59.6321	59.8443	59.6256	59.9599	59.6321
40	61.4127	61.3008	61.5338	61.3028	61.4127	61.3008	61.5338	61.3028
41	60.2172	60.0427	60.3774	60.0433	60.2172	60.0427	60.3774	60.0433
42	58.5240	58.1570	58.4004	58.1754	58.5240	58.1570	58.4004	58.1754

Figure A.1: GLCM texture features calculated from the malignant ROIs

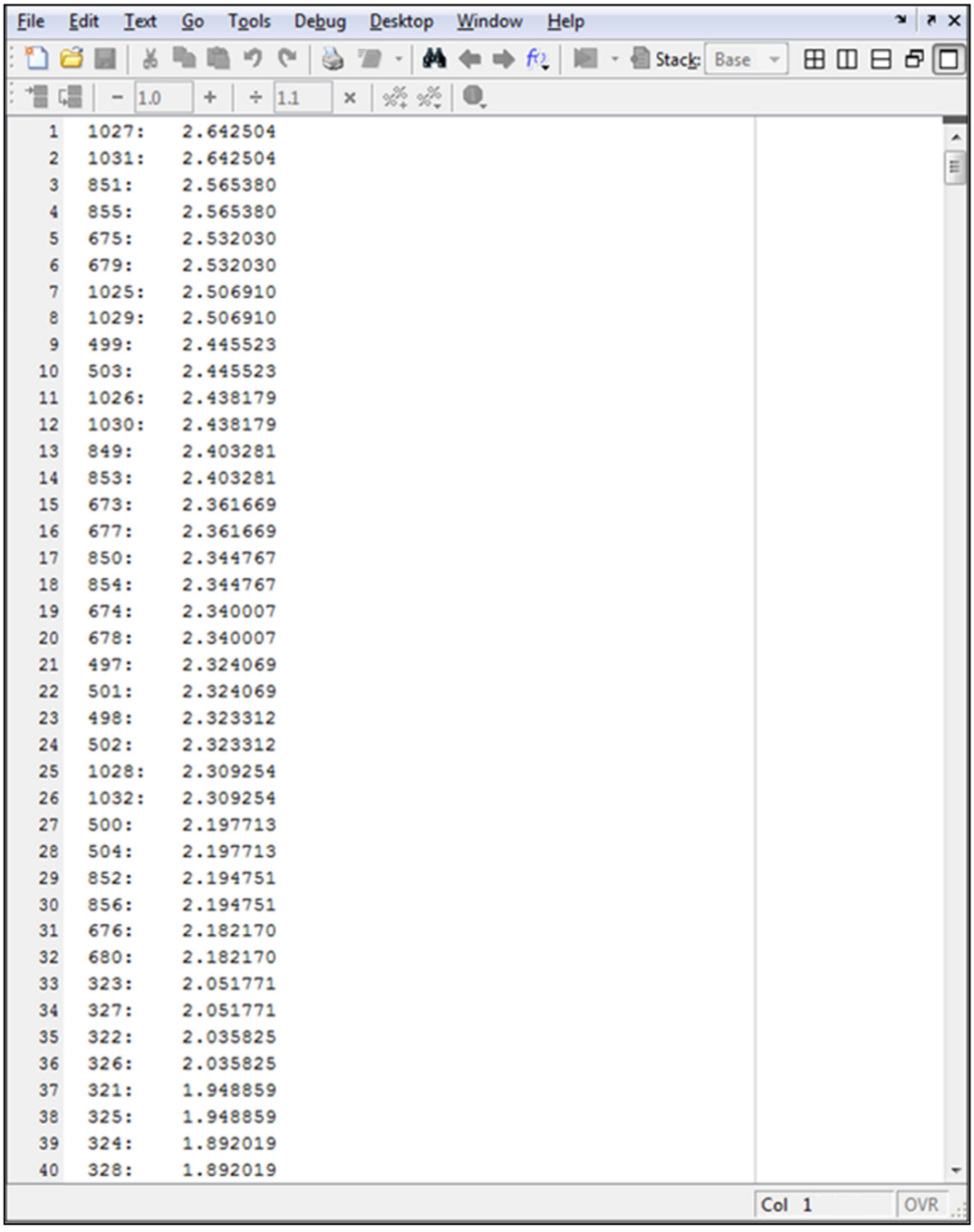
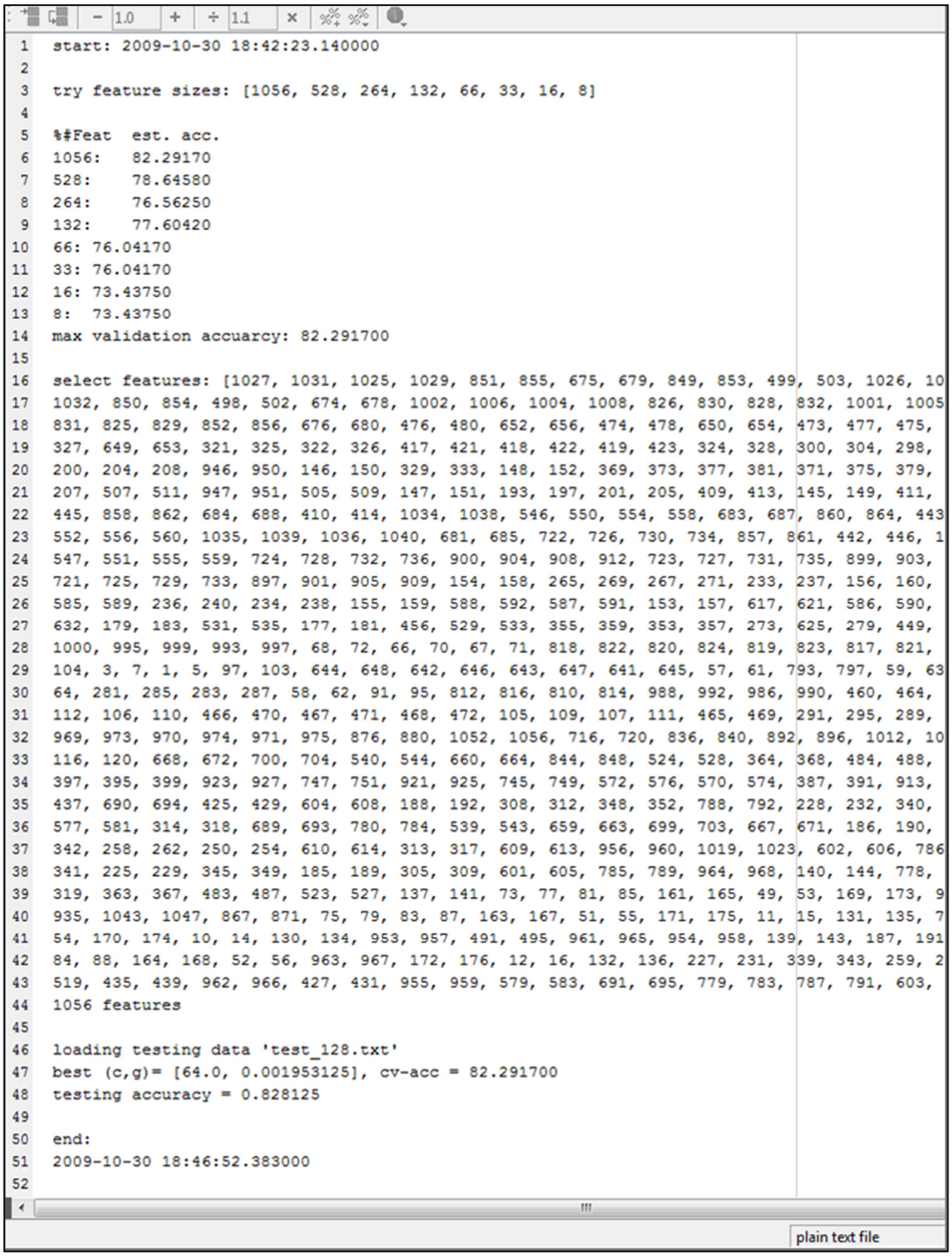


Figure A.2: F-scores of the 1152 GLCM texture feature values



```
start: 2009-10-30 18:42:23.140000

try feature sizes: [1056, 528, 264, 132, 66, 33, 16, 8]

%#Feat  est. acc.
1056:   82.29170
528:    78.64580
264:    76.56250
132:    77.60420
66: 76.04170
33: 76.04170
16: 73.43750
8:  73.43750
max validation accuarcy: 82.291700

select features: [1027, 1031, 1025, 1029, 851, 855, 675, 679, 849, 853, 499, 503, 1026, 10
1032, 850, 854, 498, 502, 674, 678, 1002, 1006, 1004, 1008, 826, 830, 828, 832, 1001, 1005
831, 825, 829, 852, 856, 676, 680, 476, 480, 652, 656, 474, 478, 650, 654, 473, 477, 475,
327, 649, 653, 321, 325, 322, 326, 417, 421, 418, 422, 419, 423, 324, 328, 300, 304, 298,
200, 204, 208, 946, 950, 146, 150, 329, 333, 148, 152, 369, 373, 377, 381, 371, 375, 379,
207, 507, 511, 947, 951, 505, 509, 147, 151, 193, 197, 201, 205, 409, 413, 145, 149, 411,
445, 858, 862, 684, 688, 410, 414, 1034, 1038, 546, 550, 554, 558, 683, 687, 860, 864, 443
552, 556, 560, 1035, 1039, 1036, 1040, 681, 685, 722, 726, 730, 734, 857, 861, 442, 446, 1
547, 551, 555, 559, 724, 728, 732, 736, 900, 904, 908, 912, 723, 727, 731, 735, 899, 903,
721, 725, 729, 733, 897, 901, 905, 909, 154, 158, 265, 269, 267, 271, 233, 237, 156, 160,
585, 589, 236, 240, 234, 238, 155, 159, 588, 592, 587, 591, 153, 157, 617, 621, 586, 590,
632, 179, 183, 531, 535, 177, 181, 456, 529, 533, 355, 359, 353, 357, 273, 625, 279, 449,
1000, 995, 999, 993, 997, 68, 72, 66, 70, 67, 71, 818, 822, 820, 824, 819, 823, 817, 821,
104, 3, 7, 1, 5, 97, 103, 644, 648, 642, 646, 643, 647, 641, 645, 57, 61, 793, 797, 59, 63
64, 281, 285, 283, 287, 58, 62, 91, 95, 812, 816, 810, 814, 988, 992, 986, 990, 460, 464,
112, 106, 110, 466, 470, 467, 471, 468, 472, 105, 109, 107, 111, 465, 469, 291, 295, 289,
969, 973, 970, 974, 971, 975, 876, 880, 1052, 1056, 716, 720, 836, 840, 892, 896, 1012, 10
116, 120, 668, 672, 700, 704, 540, 544, 660, 664, 844, 848, 524, 528, 364, 368, 484, 488,
397, 395, 399, 923, 927, 747, 751, 921, 925, 745, 749, 572, 576, 570, 574, 387, 391, 913,
437, 690, 694, 425, 429, 604, 608, 188, 192, 308, 312, 348, 352, 788, 792, 228, 232, 340,
577, 581, 314, 318, 689, 693, 780, 784, 539, 543, 659, 663, 699, 703, 667, 671, 186, 190,
342, 258, 262, 250, 254, 610, 614, 313, 317, 609, 613, 956, 960, 1019, 1023, 602, 606, 786
341, 225, 229, 345, 349, 185, 189, 305, 309, 601, 605, 785, 789, 964, 968, 140, 144, 778,
319, 363, 367, 483, 487, 523, 527, 137, 141, 73, 77, 81, 85, 161, 165, 49, 53, 169, 173, 9
935, 1043, 1047, 867, 871, 75, 79, 83, 87, 163, 167, 51, 55, 171, 175, 11, 15, 131, 135, 7
54, 170, 174, 10, 14, 130, 134, 953, 957, 491, 495, 961, 965, 954, 958, 139, 143, 187, 191
84, 88, 164, 168, 52, 56, 963, 967, 172, 176, 12, 16, 132, 136, 227, 231, 339, 343, 259, 2
519, 435, 439, 962, 966, 427, 431, 955, 959, 579, 583, 691, 695, 779, 783, 787, 791, 603,
1056 features

loading testing data 'test_128.txt'
best (c,g)= [64.0, 0.001953125], cv-acc = 82.291700
testing accuracy = 0.828125

end:
2009-10-30 18:46:52.383000
```

plain text file

Figure A.3: Optimal subset of features obtained using the "F-score + RF + SVM" technique

samples <256x264 double>

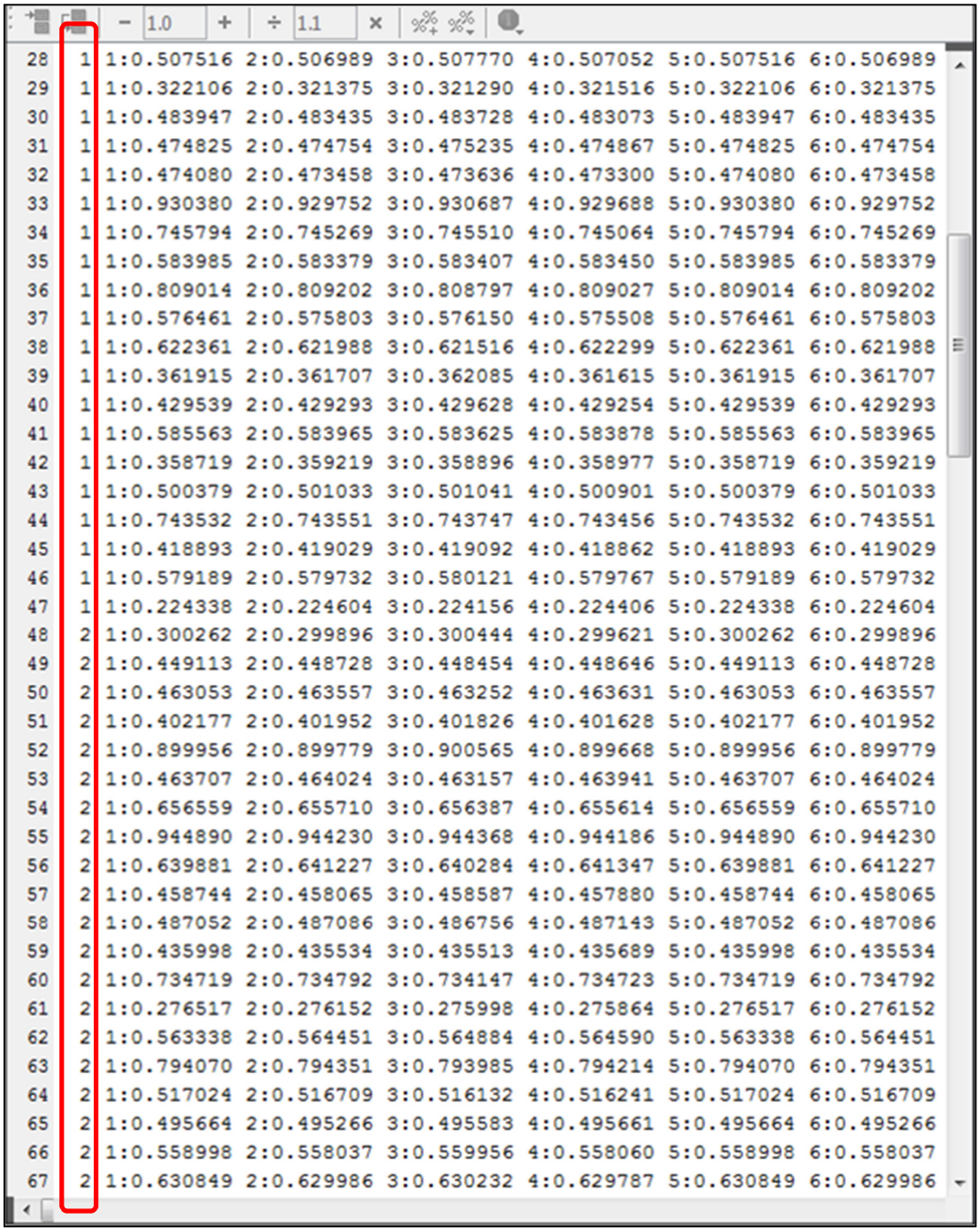
	1	2	3	4	5	6	7	8
1	-0.3154	-0.3154	-0.3764	-0.3764	-0.4009	-0.4009	-0.3962	-0.3962
2	-0.4969	-0.4969	-0.5633	-0.5633	-0.6416	-0.6416	-0.7154	-0.7154
3	-0.4935	-0.4935	-0.5325	-0.5325	-0.6032	-0.6032	-0.6698	-0.6698
4	-0.4229	-0.4229	-0.4755	-0.4755	-0.5543	-0.5543	-0.6030	-0.6030
5	-0.3611	-0.3611	-0.4073	-0.4073	-0.4809	-0.4809	-0.5403	-0.5403
6	-0.3267	-0.3267	-0.3714	-0.3714	-0.4428	-0.4428	-0.4986	-0.4986
7	-0.4397	-0.4397	-0.5032	-0.5032	-0.5752	-0.5752	-0.6478	-0.6478
8	-0.4327	-0.4327	-0.4943	-0.4943	-0.5766	-0.5766	-0.6205	-0.6205
9	-0.3369	-0.3369	-0.3891	-0.3891	-0.4467	-0.4467	-0.4824	-0.4824
10	-0.3811	-0.3811	-0.4360	-0.4360	-0.5066	-0.5066	-0.5426	-0.5426
11	-0.4344	-0.4344	-0.4800	-0.4800	-0.5477	-0.5477	-0.5868	-0.5868
12	-0.3621	-0.3621	-0.4183	-0.4183	-0.4862	-0.4862	-0.5494	-0.5494
13	-0.3484	-0.3484	-0.3992	-0.3992	-0.4686	-0.4686	-0.5229	-0.5229
14	-0.4513	-0.4513	-0.5036	-0.5036	-0.5817	-0.5817	-0.6407	-0.6407
15	-0.4289	-0.4289	-0.4887	-0.4887	-0.5649	-0.5649	-0.6063	-0.6063
16	-0.4317	-0.4317	-0.4768	-0.4768	-0.5553	-0.5553	-0.6128	-0.6128
17	-0.4969	-0.4969	-0.5160	-0.5160	-0.5817	-0.5817	-0.6523	-0.6523
18	-0.3919	-0.3919	-0.4549	-0.4549	-0.5065	-0.5065	-0.4731	-0.4731
19	-0.4502	-0.4502	-0.4931	-0.4931	-0.5716	-0.5716	-0.6384	-0.6384
20	-0.3768	-0.3768	-0.4297	-0.4297	-0.5120	-0.5120	-0.5710	-0.5710
21	-0.3484	-0.3484	-0.4021	-0.4021	-0.4766	-0.4766	-0.5109	-0.5109
22	-0.3924	-0.3924	-0.4467	-0.4467	-0.5270	-0.5270	-0.5949	-0.5949
23	-0.2943	-0.2943	-0.3412	-0.3412	-0.3975	-0.3975	-0.4348	-0.4348
24	-0.4120	-0.4120	-0.4824	-0.4824	-0.5531	-0.5531	-0.5661	-0.5661
25	-0.4066	-0.4066	-0.4602	-0.4602	-0.5318	-0.5318	-0.5727	-0.5727
26	-0.4795	-0.4795	-0.5192	-0.5192	-0.5944	-0.5944	-0.6637	-0.6637
27	-0.4436	-0.4436	-0.4846	-0.4846	-0.5537	-0.5537	-0.6254	-0.6254
28	-0.3947	-0.3947	-0.4436	-0.4436	-0.5158	-0.5158	-0.5822	-0.5822
29	-0.3723	-0.3723	-0.4178	-0.4178	-0.4852	-0.4852	-0.5436	-0.5436
30	-0.3508	-0.3508	-0.4004	-0.4004	-0.4706	-0.4706	-0.5335	-0.5335
31	-0.4428	-0.4428	-0.4811	-0.4811	-0.5572	-0.5572	-0.6222	-0.6222
32	-0.3688	-0.3688	-0.4145	-0.4145	-0.4877	-0.4877	-0.5450	-0.5450
33	-0.2324	-0.2324	-0.2733	-0.2733	-0.3265	-0.3265	-0.3083	-0.3083
34	-0.3214	-0.3214	-0.3611	-0.3611	-0.4336	-0.4336	-0.4791	-0.4791
35	-0.4188	-0.4188	-0.4564	-0.4564	-0.5335	-0.5335	-0.5963	-0.5963
36	-0.3551	-0.3551	-0.4086	-0.4086	-0.4673	-0.4673	-0.5107	-0.5107
37	-0.2245	-0.2245	-0.2551	-0.2551	-0.3054	-0.3054	-0.3322	-0.3322
38	-0.3704	-0.3704	-0.4068	-0.4068	-0.4769	-0.4769	-0.5377	-0.5377
39	-0.4204	-0.4204	-0.4552	-0.4552	-0.5256	-0.5256	-0.6058	-0.6058
40	-0.2952	-0.2952	-0.3386	-0.3386	-0.4115	-0.4115	-0.4460	-0.4460
41	-0.2939	-0.2939	-0.3500	-0.3500	-0.3990	-0.3990	-0.3978	-0.3978
42	-0.4026	-0.4026	-0.4468	-0.4468	-0.5211	-0.5211	-0.5764	-0.5764

Figure A.4: Texture features values for optimum subset of 1056 features

samples2 <256x264 double>

	1	2	3	4	5	6	7	8
1	0.5114	0.5114	0.4444	0.4444	0.4772	0.4772	0.5258	0.5258
2	0.1453	0.1453	0.0898	0.0898	0.0716	0.0716	0.0277	0.0277
3	0.1522	0.1522	0.1483	0.1483	0.1363	0.1363	0.0988	0.0988
4	0.2946	0.2946	0.2564	0.2564	0.2186	0.2186	0.2030	0.2030
5	0.4192	0.4192	0.3857	0.3857	0.3424	0.3424	0.3008	0.3008
6	0.4885	0.4885	0.4539	0.4539	0.4066	0.4066	0.3660	0.3660
7	0.2608	0.2608	0.2038	0.2038	0.1834	0.1834	0.1332	0.1332
8	0.2749	0.2749	0.2208	0.2208	0.1810	0.1810	0.1757	0.1757
9	0.4681	0.4681	0.4204	0.4204	0.4001	0.4001	0.3913	0.3913
10	0.3789	0.3789	0.3313	0.3313	0.2990	0.2990	0.2973	0.2973
11	0.2715	0.2715	0.2478	0.2478	0.2297	0.2297	0.2284	0.2284
12	0.4172	0.4172	0.3649	0.3649	0.3335	0.3335	0.2867	0.2867
13	0.4448	0.4448	0.4011	0.4011	0.3630	0.3630	0.3280	0.3280
14	0.2374	0.2374	0.2031	0.2031	0.1724	0.1724	0.1442	0.1442
15	0.2825	0.2825	0.2314	0.2314	0.2008	0.2008	0.1978	0.1978
16	0.2768	0.2768	0.2540	0.2540	0.2169	0.2169	0.1878	0.1878
17	0.1454	0.1454	0.1796	0.1796	0.1725	0.1725	0.1261	0.1261
18	0.3571	0.3571	0.2954	0.2954	0.2992	0.2992	0.4058	0.4058
19	0.2396	0.2396	0.2230	0.2230	0.1895	0.1895	0.1479	0.1479
20	0.3875	0.3875	0.3433	0.3433	0.2899	0.2899	0.2530	0.2530
21	0.4448	0.4448	0.3956	0.3956	0.3495	0.3495	0.3468	0.3468
22	0.3562	0.3562	0.3110	0.3110	0.2646	0.2646	0.2157	0.2157
23	0.5540	0.5540	0.5112	0.5112	0.4830	0.4830	0.4655	0.4655
24	0.3167	0.3167	0.2433	0.2433	0.2206	0.2206	0.2607	0.2607
25	0.3275	0.3275	0.2855	0.2855	0.2566	0.2566	0.2503	0.2503
26	0.1805	0.1805	0.1735	0.1735	0.1511	0.1511	0.1084	0.1084
27	0.2529	0.2529	0.2391	0.2391	0.2197	0.2197	0.1680	0.1680
28	0.3515	0.3515	0.3169	0.3169	0.2836	0.2836	0.2355	0.2355
29	0.3966	0.3966	0.3658	0.3658	0.3352	0.3352	0.2958	0.2958
30	0.4401	0.4401	0.3989	0.3989	0.3597	0.3597	0.3115	0.3115
31	0.2545	0.2545	0.2458	0.2458	0.2138	0.2138	0.1730	0.1730
32	0.4036	0.4036	0.3722	0.3722	0.3310	0.3310	0.2936	0.2936
33	0.6788	0.6788	0.6402	0.6402	0.6027	0.6027	0.6630	0.6630
34	0.4993	0.4993	0.4736	0.4736	0.4221	0.4221	0.3964	0.3964
35	0.3028	0.3028	0.2927	0.2927	0.2536	0.2536	0.2135	0.2135
36	0.4314	0.4314	0.3834	0.3834	0.3653	0.3653	0.3471	0.3471
37	0.6947	0.6947	0.6746	0.6746	0.6381	0.6381	0.6256	0.6256
38	0.4005	0.4005	0.3867	0.3867	0.3492	0.3492	0.3049	0.3049
39	0.2997	0.2997	0.2949	0.2949	0.2671	0.2671	0.1986	0.1986
40	0.5521	0.5521	0.5161	0.5161	0.4594	0.4594	0.4481	0.4481
41	0.5547	0.5547	0.4945	0.4945	0.4805	0.4805	0.5233	0.5233
42	0.3355	0.3355	0.3109	0.3109	0.2747	0.2747	0.2446	0.2446

Figure A.5: Normalized feature values in the range between 0 and 1



28	1	1:0.507516	2:0.506989	3:0.507770	4:0.507052	5:0.507516	6:0.506989
29	1	1:0.322106	2:0.321375	3:0.321290	4:0.321516	5:0.322106	6:0.321375
30	1	1:0.483947	2:0.483435	3:0.483728	4:0.483073	5:0.483947	6:0.483435
31	1	1:0.474825	2:0.474754	3:0.475235	4:0.474867	5:0.474825	6:0.474754
32	1	1:0.474080	2:0.473458	3:0.473636	4:0.473300	5:0.474080	6:0.473458
33	1	1:0.930380	2:0.929752	3:0.930687	4:0.929688	5:0.930380	6:0.929752
34	1	1:0.745794	2:0.745269	3:0.745510	4:0.745064	5:0.745794	6:0.745269
35	1	1:0.583985	2:0.583379	3:0.583407	4:0.583450	5:0.583985	6:0.583379
36	1	1:0.809014	2:0.809202	3:0.808797	4:0.809027	5:0.809014	6:0.809202
37	1	1:0.576461	2:0.575803	3:0.576150	4:0.575508	5:0.576461	6:0.575803
38	1	1:0.622361	2:0.621988	3:0.621516	4:0.622299	5:0.622361	6:0.621988
39	1	1:0.361915	2:0.361707	3:0.362085	4:0.361615	5:0.361915	6:0.361707
40	1	1:0.429539	2:0.429293	3:0.429628	4:0.429254	5:0.429539	6:0.429293
41	1	1:0.585563	2:0.583965	3:0.583625	4:0.583878	5:0.585563	6:0.583965
42	1	1:0.358719	2:0.359219	3:0.358896	4:0.358977	5:0.358719	6:0.359219
43	1	1:0.500379	2:0.501033	3:0.501041	4:0.500901	5:0.500379	6:0.501033
44	1	1:0.743532	2:0.743551	3:0.743747	4:0.743456	5:0.743532	6:0.743551
45	1	1:0.418893	2:0.419029	3:0.419092	4:0.418862	5:0.418893	6:0.419029
46	1	1:0.579189	2:0.579732	3:0.580121	4:0.579767	5:0.579189	6:0.579732
47	1	1:0.224338	2:0.224604	3:0.224156	4:0.224406	5:0.224338	6:0.224604
48	2	1:0.300262	2:0.299896	3:0.300444	4:0.299621	5:0.300262	6:0.299896
49	2	1:0.449113	2:0.448728	3:0.448454	4:0.448646	5:0.449113	6:0.448728
50	2	1:0.463053	2:0.463557	3:0.463252	4:0.463631	5:0.463053	6:0.463557
51	2	1:0.402177	2:0.401952	3:0.401826	4:0.401628	5:0.402177	6:0.401952
52	2	1:0.899956	2:0.899779	3:0.900565	4:0.899668	5:0.899956	6:0.899779
53	2	1:0.463707	2:0.464024	3:0.463157	4:0.463941	5:0.463707	6:0.464024
54	2	1:0.656559	2:0.655710	3:0.656387	4:0.655614	5:0.656559	6:0.655710
55	2	1:0.944890	2:0.944230	3:0.944368	4:0.944186	5:0.944890	6:0.944230
56	2	1:0.639881	2:0.641227	3:0.640284	4:0.641347	5:0.639881	6:0.641227
57	2	1:0.458744	2:0.458065	3:0.458587	4:0.457880	5:0.458744	6:0.458065
58	2	1:0.487052	2:0.487086	3:0.486756	4:0.487143	5:0.487052	6:0.487086
59	2	1:0.435998	2:0.435534	3:0.435513	4:0.435689	5:0.435998	6:0.435534
60	2	1:0.734719	2:0.734792	3:0.734147	4:0.734723	5:0.734719	6:0.734792
61	2	1:0.276517	2:0.276152	3:0.275998	4:0.275864	5:0.276517	6:0.276152
62	2	1:0.563338	2:0.564451	3:0.564884	4:0.564590	5:0.563338	6:0.564451
63	2	1:0.794070	2:0.794351	3:0.793985	4:0.794214	5:0.794070	6:0.794351
64	2	1:0.517024	2:0.516709	3:0.516132	4:0.516241	5:0.517024	6:0.516709
65	2	1:0.495664	2:0.495266	3:0.495583	4:0.495661	5:0.495664	6:0.495266
66	2	1:0.558998	2:0.558037	3:0.559956	4:0.558060	5:0.558998	6:0.558037
67	2	1:0.630849	2:0.629986	3:0.630232	4:0.629787	5:0.630849	6:0.629986

Figure A.6: Training data feature file with malignant and benign class labels

APPENDIX B

SVM Training and Testing

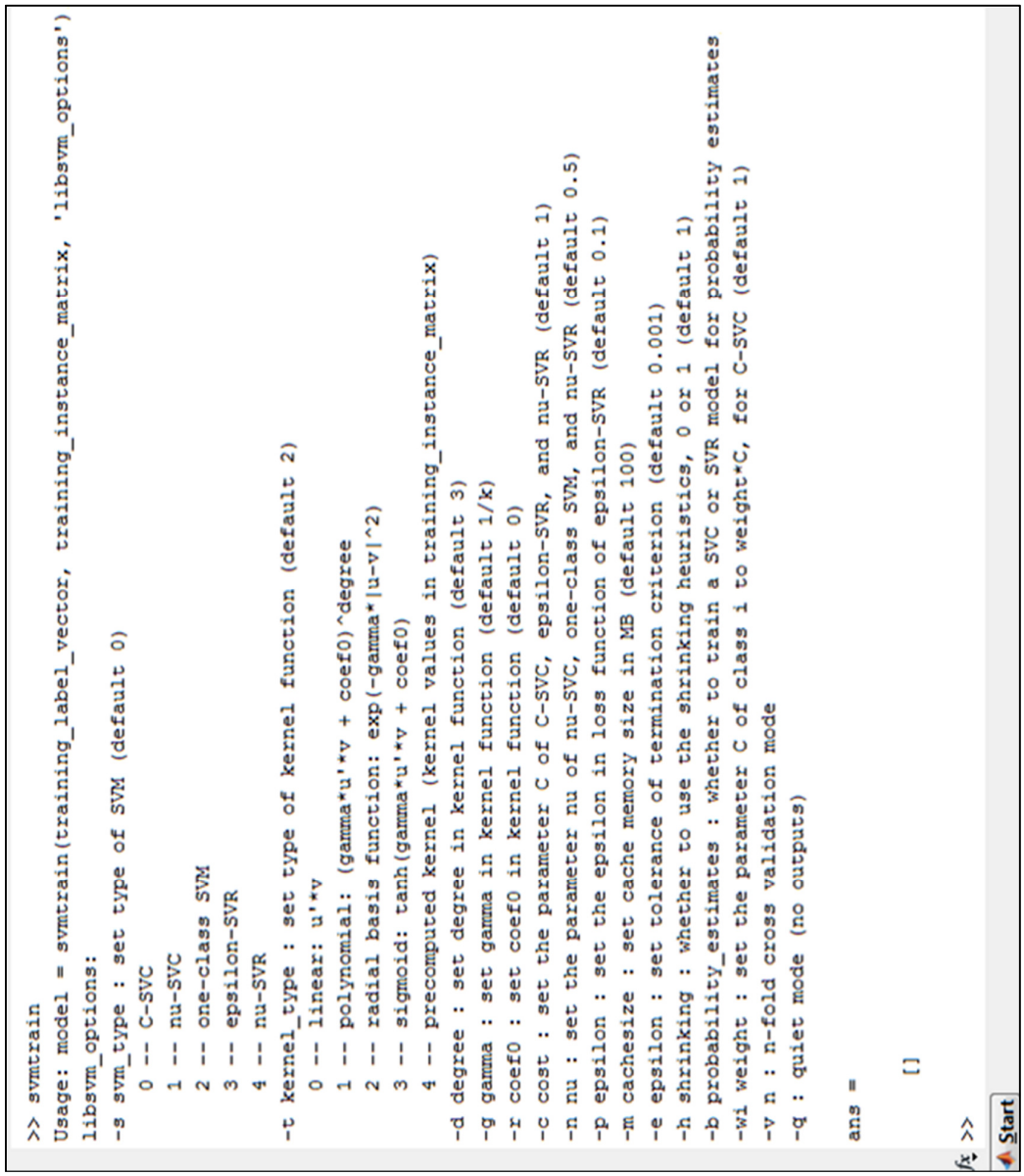


Figure B.1: The LIBSVM SVM training function in MATLAB

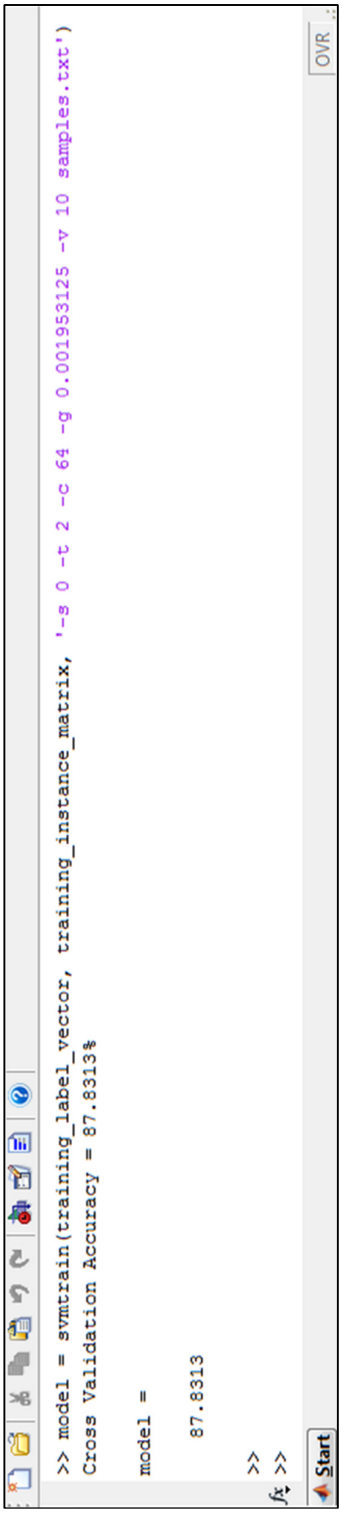


Figure B.2: SVM model parameters in MATLAB after SVM training

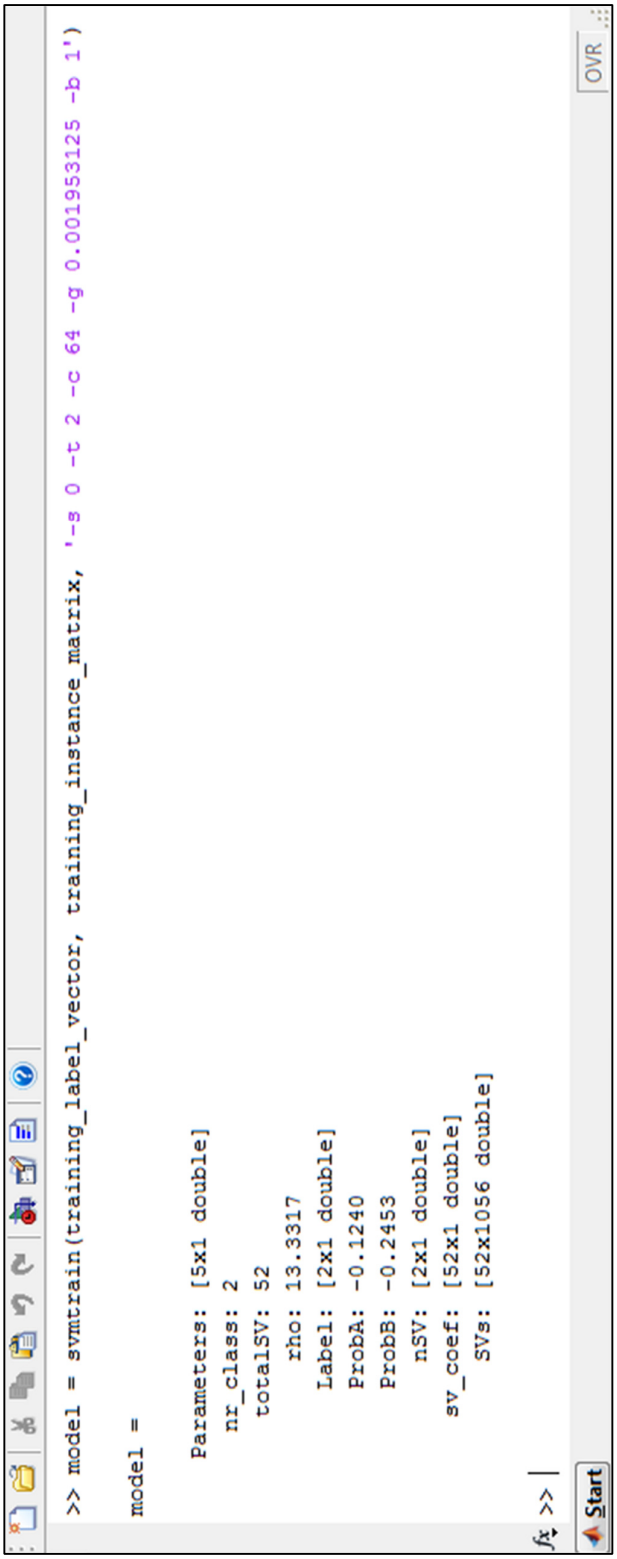


Figure B.3: SVM model generated after SVM training in MATLAB

```
>> svmpredict
Usage: [predicted_label, accuracy, decision_values/prob_estimates] = svmpredict(testing_label_vector, testing_instance_matrix, model, 'libsvm_options')
Parameters:
  model: SVM model structure from svmtrain.
  libsvm_options:
    -b probability_estimates: whether to predict probability estimates, 0 or 1 (default 0); one-class SVM not supported yet
Returns:
  predicted_label: SVM prediction output vector.
  accuracy: a vector with accuracy, mean squared error, squared correlation coefficient.
  prob_estimates: If selected, probability estimate vector.

ans =

     []

>>
```

Figure B.4: The LIBSVM prediction function in MATLAB

Figure B.5: SVM testing and validation in MATLAB using LIBSVM

APPENDIX C

LIBSVM Copyright Notice

The breast cancer detection system developed in this research is collaborated with the Faculty of Computer Science and Information Technology (FSKTM), University of Malaya and the Department of Radiology, University of Malaya Medical Centre (UMMC), Kuala Lumpur. The system developed incorporates a tool "LIBSVM", a library for support vector machines, which was developed by Chih-Chung Chang and Chih-Jen Lin. Acknowledgement of the LIBSVM copyright is shown as below.

APPENDIX D

List of Publications

Journal Publications

[1] Jawad Nagi, Keem Siah Yap, Sieh Kiong Tiong, Syed Khaleel Ahmed, Farrukh Nagi, ***"Improving SVM-based Nontechnical Loss Detection in Power Utility Using Fuzzy Inference System"***, accepted for publication in *IEEE Transactions on Power Delivery* on 22nd June 2010. Manuscript ID: PESL-00108-2009.R2.

[2] Mohammad Mehdi Badjian, Jawad Nagi, Sieh Kiong Tiong, Keem Siah Yap, Siaw Paw Koh, Farrukh Nagi, ***"Comparison of Supervised Learning Techniques for Non-Technical Loss Detection in Power Utility"***, submitted to *Malaysian Journal of Computer Science* (MJCS) for first review on 9 April 2010.

[3] Farrukh Nagi, Syed Khaleel Ahmed, Jawad Nagi, ***"Fuzzy Time-Optimal Controller (FTOC) for Second Order Nonlinear Systems"***, submitted to *IEEE Transactions on Systems, Man, and Cybernetics: Part B* for first review on 30 November 2009. Paper No: SMCB-E-2009-11-1046.

[4] Jawad Nagi, Keem Siah Yap, Sieh Kiong Tiong, Syed Khaleel Ahmed, Farrukh Nagi, ***"A Computational Intelligence Scheme for Prediction of the Daily Peak Load"***, submitted for second review to *Applied Soft Computing* (ASOC) on 10 August 2010. Manuscript Reference No: ASOC-D-09-00556.

[5] Jawad Nagi, Keem Siah Yap, Sieh Kiong Tiong, Syed Khaleel Ahmed, and Malik Mohammad, **"*Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines*"**, *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 1162–1171, Apr. 2010.

[6] Farrukh Nagi, Logah Perumal, and Jawad Nagi, **"*A New Integrated Fuzzy Bang-Bang Relay Control System*"**, *Mechatronics*, vol. 19, no. 5, pp. 748–760, Aug. 2009.

Conference Publications

[1] Jawad Nagi, Keem Siah Yap, Farrukh Nagi, Sieh Kiong Tiong, Siaw Paw Koh, Syed Khaleel Ahmed, **"*NTL Detection of Electricity Theft and Abnormalities for Large Power Consumers in TNB Malaysia*"**, in Proc. of the 2010 IEEE Student Conference on Research and Development (SCOReD) 2010, 14 Dec. 2010, Malaysia, pp. 1–5.

[2] Jawad Nagi, Sameem Abdul Kareem, Farrukh Nagi, and Syed Khaleel Ahmed, "***Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms***", in Proc. of the IEEE Conference on Biomedical Engineering and Sciences (IECBES) 2010, 30 Nov. 2010, Kuala Lumpur, Malaysia, pp. 1–6.

[3] Jawad Nagi, Tiong Sieh Kiong, Syed Khaleel Ahmed, and Farrukh Nagi, ***"Prediction of PVT Properties in Crude Oil Systems Using Support Vector Machines",*** in Proc. of the 3rd International Conference on Energy and Environment (ICEE) 2009, Dec. 7-8, 2009, Malacca, Malaysia, pp. 1–5.

[4] Jawad Nagi, Keem Siah Yap, Sieh Kiong Tiong, Abdul Malik Mohammad, and Syed Khaleel Ahmed, **"*Non-Technical Loss Analysis for Detection of Electricity Theft using Support Vector Machines*"**, in Proc. of the 2nd IEEE International Power and Energy Conference (PECon) 2008, Dec. 1-3, 2008, Johor Bahru, Malaysia, pp. 907–912.

[5] Jawad Nagi, Keem Siah Yap, Sieh Kiong Tiong, and Syed Khaleel Ahmed, **"*Detection of Abnormalities and Electricity Theft using Genetic Support Vector Machines*"**, in Proc. of the IEEE Region 10 Conference (TENCON) 2008, Nov. 19, 2008, Hyderabad, India, pp. 1–6.

[6] Jawad Nagi, Syed Khaleel Ahmed, and Farrukh Nagi, **"*Pose Invariant Face Recognition using Hybrid DWT-DCT Frequency Features with Support Vector Machines*"**, in Proc. of the 4th International Conference on Information Technology and Multimedia at UNITEN (ICIMu) 2008, Nov. 18-19, 2008, Bandar Baru Bangi, Selangor, Malaysia, pp. 99–104.

[7] Jawad Nagi, Keem Siah Yap, Sieh Kiong Tiong, and Abdul Malik Mohammad, **"*Intelligent System for Detection of Abnormalities and Theft of Electricity using Genetic Algorithm and Support Vector Machines*"**, in Proc. of the 4th International Conference on Information Technology and Multimedia at UNITEN (ICIMu) 2008, Nov. 18-19, 2008, Bandar Baru Bangi, Selangor, Malaysia, pp. 122–127.

[8] Jawad Nagi, Keem Siah Yap, Sieh Kiong Tiong, Syed Khaleel Ahmed, and Farrukh Nagi, ***"Intelligent Detection of DTMF Tones using a Hybrid Signal Processing Technique with Support Vector Machines",*** in Proc. of the International Symposium on Information Technology (ITSIM) 2008, Aug. 26-28, 2008, Kuala Lumpur, Malaysia, vol. 4, pp. 1–8.

[9] Jawad Nagi, Sieh Kiong Tiong, Yap Keem Siah, and Syed Khaleel Ahmed, ***"Dual-tone Multi-frequency Signal Detection using Support Vector Machines"***, in Proc. of the 6th National Conference on Telecommunication Technologies and Malaysia Conference on Photonics (NCTT-MCP) 2008, Aug. 26-28, 2008, Putrajaya, Malaysia, pp. 350–355.

[10] Jawad Nagi, Syed Khaleel Ahmed, and Farrukh Nagi, ***"Palm Biodiesel an Alternative Green Renewable Energy for the Energy Demands of the Future"***, in Proc. of the International Conference on Renewable Energy and Sustainability (ICCBT) 2008, Jun. 16-20, 2008, Kuala Lumpur, Malaysia, pp. 79–94.

[11] Jawad Nagi, Keem Siah Yap, Sieh Kiong Tiong, and Syed Khaleel Ahmed, ***"Electrical Power Load Forecasting using Hybrid Self-Organizing Maps and Support Vector Machines"***, in Proc. of the 2nd International Power Engineering and Optimization Conference (PEOCO) 2008, Jun. 4-5, 2008, Shah Alam, Malaysia, pp. 51–56.

[12] Jawad Nagi, Syed Khaleel Ahmed, and Farrukh Nagi, ***"A MATLAB based Face Recognition System using Image Processing and Neural Networks"***, in Proc. of the 4th International Colloquium on Signal Processing and its Applications (CSPA) 2008, Mar. 7-9, 2008, Kuala Lumpur, Malaysia, pp. 83–88.

BIODATA OF THE AUTHOR

Jawad Nagi, was born in Karachi, Pakistan on March 23, 1985. He received his Bachelor's degree from Universiti Tenaga Nasional (UNITEN), Malaysia with Honors in Electrical and Electronics Engineering in 2007. In 2009 he was awarded the Master of Electrical Engineering degree from UNITEN in 2009. He is currently pursuing a Master's of Computer Science degree at University of Malaya, Malaysia, which is expected to complete in August 2010.

He is currently working as a Research Engineer at UNITEN R&D Sdn. Bhd. of Universiti Tenaga Nasional (UNITEN) since January 2008. He also is involved in teaching activities at the Asia Pacific Institute of Information Technology (APIIT), Kuala Lumpur, Malaysia.

His research interests include pattern recognition, machine learning, image processing, load forecasting, fuzzy logic, neural networks, support vector machines, robotics and control systems. His publications and resume can be found at: **http://metalab.uniten.edu.my/~jawad/papers/**