

Computer Engineering Department
Faculty of Engineering
Deanery of Higher Studies
Islamic University – Gaza
Palestine



Mammogram Computer- Aided Diagnosis

Belal Kamal Elfarra

Supervisor

Prof. Ibrahim S. I. Abuhaiba

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Computer Engineering

1433H (2012)

ACKNOWLEDGEMENTS

I would like to thank my parents for providing me with the opportunity to be where I am. Without them, none of this would be even possible to do. You have always been around supporting and encouraging me and I appreciate that. I would also like to thank my brothers and sisters for their encouragement, input and constructive criticism which are really priceless. An extra special thanks goes to my sister Amal who did not spare any effort to review and audit my thesis linguistically.

My heartiest gratitude to my wonderful wife, Maram, for her patience and forbearance through my studying and preparing this study.

I'm deeply grateful for my supervisor, Prof. Ibrahim Abuhaiba for his serious supervision and continuous follow-up of my research. Really I've benefited a lot from his experience, accuracy and his time management.

Also I would like to thank all my teachers in master degree, for their perceptiveness, understanding, and their skillful teaching. And many thanks to all my friends for their moral support.

And above all, many thanks to Allah, who helps me get all required resources for completing this research.

CONTENTS

ACKNOWLEDGEMENTS.....	iii
LIST OF ABBREVIATION	vi
LIST OF FIGURES.....	vii
ABSTRACT	x
1. Chapter 1: Introduction.....	1
1.1 Background.....	1
1.2 Research Overview	5
1.2.1 Motivation.....	5
1.2.2 Objectives	6
1.2.3 Methodology	6
1.2.4 Contributions.....	6
1.2.5 Organization of This Thesis	7
2. Chapter 2: Related Work	8
2.1 Segmentation.....	8
2.2 Feature Extraction and Selection	9
2.3 Classification.....	12
3. Chapter 3: Background	14
3.1 CADx System	14
3.1.1 Preprocessing	14
3.1.2 Segmentation.....	16
3.1.3 Feature Extraction	17
3.1.4 Feature Selection.....	17
3.1.5 Classification.....	20
3.2 Performance Metrics.....	23
4. Chapter 4: Proposed Feature Extraction and Selection Techniques	26
4.1 Introduction.....	26
4.2 Feature Extraction.....	26
4.2.1 Spatial Gray Level Dependence Method (SGLDM).....	27
4.2.2 Run Difference Matrix (RDM)	32
4.2.3 Square Centroid Lines Gray Level Distribution Method (SCLGM).....	36
4.2.4 Human Features	47
4.3 Feature Selection.....	48
5. Chapter 5: Experimentation and Results	52

5.1	DDSM Database	52
5.2	Converting LJPEG to TIF	53
5.3	Implementation Environment	54
5.4	Manual Segmentation	54
5.5	Enhancement.....	55
5.6	Segmentation.....	56
5.7	Feature Extraction.....	58
5.8	Feature Selection.....	59
5.9	Classification.....	62
5.10	Results Discussion	72
6.	Chapter 6: Conclusion and Future Work.....	75
6.1	Summary and Concluding Remarks	75
6.2	Future Work.....	77
	Bibliography	78

LIST OF ABBREVIATION

ACR	: American College Of Radiology
ANN	: Artificial Neural Network
BI-RADS	: Breast Imaging Reporting And Data System
BWC	: Benign Without Callback
CADe	: Computer Aided Detection
CADx	: Computer Aided Diagnosis
CC	: CranioCaudal
CDF	: Cumulative Density Function
DAD	: Distribution of Average Distance
DDSM	: Digital Database for Screening Mammography
DF	: Desired Number
DGD	: Distribution of Gray level Difference
DOD	: Distribution Of average Difference
FF-ANN	: Feed-Forward Artificial Neural Network
FPP	: False Positive Prediction
FS	: Feature Selection
GA	: Genetic Algorithm
GAF	: Genetic Algorithm Features
GLCM	: Gray Level Co-occurrence Matrix
ISF	: Intersection of Selected Features
K	: Kurtosis
KL	: Karhunen-Loeve transform
LDA	: linear discriminant analysis
LDE	: Large Difference Emphasis
LDEL	: Long Distance Emphasis for Large difference
LJPEG	: Lossless JPEG
MAD	: Mean Absolute Deviation
MLO	: MedioLateral Oblique
mRMR	: Max-Relevance and Min-Redundancy
NN	: Neural network
PDF	: Probability Density Function
RDM	: Run Difference Method
ROC	: Receiver Operating Characteristics curve
ROI	: Region Of Interest
SCLGM	: Square Centroid Lines Gray level distribution Method
SFF	: Sequential Forward Features
SGLDM	: Spatial Gray Level Dependence Method
SK	: SKewness
SMG	: Second Moment of DGD
SMO	: Second Moment Of DOD
SVM	: Support Vector Machine
TIF	: Tag Image File Format
USF	: Union of Selected Features

LIST OF FIGURES

Figure 1-1: In mammogram two views are taken for each breast.....	2
Figure 1-2: An example of a rounded circumscribed mass, this image is the case “A_0364 left” obtained from Digital Database for Screening Mammography DDSM.....	2
Figure 1-3: An example of microcalcification shown as small bright spots, this image is the case “A_1115 Right” obtained from Digital Database for Screening Mammography DDSM..	3
Figure 1-4: An example of a mass with Architectural distortions shape, this image is the case “A_1478 left” obtained from Digital Database for Screening Mammography DDSM.	3
Figure 1-5: Some of mass shapes.....	4
Figure 1-6: Mass margins.	4
Figure 3-1: Mammogram Computer-Aided Diagnosis (CADx) steps.	15
Figure 3-2: A basic genetic algorithm is a stochastic iterative search method [34].....	19
Figure 3-3: SVM will choose the hyperpalne H2 by which we have best margin [35].	21
Figure 3-4: Samples that located on the margin called the support vector [35].....	21
Figure 3-5: the structure of one neurons in ANN	23
Figure 3-6: An example of three-layer of ANN [35].	24
Figure 3-7: ROC curve.....	25
Figure 4-1: The co-occurrence for one pixel (yellow pixel) with $d=3$ pixel and $\theta \in \{0, \pi/4, 2\pi/4, 3\pi/4\}$	27
Figure 4-2: Two example of co-occurrence matrix with $d=1$ and $\theta=0^\circ$ (upper) and $\theta=45^\circ$ (lower).....	28
Figure 4-3: Run Difference Matrix (RDM) is a function of r and level difference with the given direction θ	33
Figure 4-4: Centroid lines pixels provide discriminative information about the mass type. ...	36
Figure 4-5: The mutual information $I(x;y)$ is the intersection of the entropies of each x and y	50
Figure 4-6: Feature selection technique.	50
Figure 5-1: An example for each normal, benign and cancer.	53
Figure 5-2: Some examples of manual segmentation.	55
Figure 5-3: Histogram of original mammogram (left) and after enhancement (right).....	56
Figure 5-4: Mammogram images enhancements.	56
Figure 5-5: By multiplying the enhanced image (first raw) with the mask (second raw) we obtain the segmented mass shown in the third raw.....	57
Figure 5-6: Use SVM for three classes	60
Figure 5-7: ROC curve for FANN classifier using SFF features.....	65

Figure 5-8: Confusion matrix for FANN classifier using SFF features.....	65
Figure 5-9: Confusion matrix and ROC curve of testing dataset-1 using SFF features.....	65
Figure 5-10: Confusion matrix and ROC curve of testing dataset-2 using SFF features.....	66
Figure 5-11: ROC curve for FANN classifier using GAF features.	67
Figure 5-12: Confusion matrix for FF-ANN classifier using GAF features.....	67
Figure 5-13: Confusion matrix and ROC curve of testing dataset-1 using GAF features.	68
Figure 5-14: Confusion matrix and ROC curve of testing dataset-2 using GAF features.	68
Figure 5-15: The ROC curve of the classifier FF-ANN using ISF features.	69
Figure 5-16: Confusion matrix for FF-ANN classifier using ISF features.	69
Figure 5-17: With shared features, the best validation performance is 0.062948 at epoch 14.70	
Figure 5-18: Confusion matrix and ROC curve of testing dataset-1 using ISF features.	70
Figure 5-19: Confusion matrix and ROC curve of testing dataset-2 using ISF features.	70
Figure 5-20: ROC curve for the classifier using USF features.	71
Figure 5-21: Confusion matrix for FF-ANN classifier using USF features.....	72
Figure 5-22: With USF features, the best validation performance is 0.030551 at epoch 40. ..	72
Figure 5-23: Confusion matrix and ROC curve of testing dataset-1 using USF features.	73
Figure 5-24: Confusion matrix and ROC curve of testing dataset-2 using USF features.	73
Figure 5-25: Examples of misclassified mammogram images	74

تشخيص أورام الثدي بمساعدة الحاسوب

بلال كمال الفرا

ملخص

نظام "تشخيص أورام الثدي بمساعدة الحاسوب" يستخدم لمساعدة الأطباء في تشخيص وتحديد نوع الأورام المكتشفة في صورة الماموجرام، والماموجرام هو فحص للثدي يستخدم الأشعة السينية ذات طاقة إشعاعية محدودة لها القدرة على اظهار الورم بوضوح. تحسين وتطوير نظام التشخيص المحوسب قد يؤدي الى إكتشاف الورم قبل الإحساس به من قبل المريضة بفترة تصل إلى سنتين. الهدف الرئيسي من هذا البحث هو تحسين نظام التشخيص المحوسب من خلال تحسين وإبتكار طرق إستنباط صفات الأورام في صورة الأشعة وتحسين طرق إختيار الأفضل من هذه الصفات. تم استخدام نوعين من الصفات في هذا البحث. النوع الأول هو النوع الملاحظ بالعين البشرية وتم الحصول على صفات هذا النوع من قواعد البيانات الرقمية للماموجرام بالأشعة السينية من جامعة فلوريدا. والنوع الآخر هو النوع الحسابي وتم إستنباط صفات هذا النوع من خلال ثلاث طرق: 1- طريقة الفروقات بين قيم نقاط متباعدة بمسافة وزاوية محددة، 2- طريقة علاقة كل نقطة بما يجاورها من النقاط حسب درجة الرمادية وتم التعديل على الطريقتين المذكورتين بهدف الحصول على نتائج أقوى 3- طريقة الخطوط المركزية لأصغر مربع يحتوي على الورم والعلاقة فيما بينهم وهي طريقه جديده نقدمها من خلال هذا البحث. بعد إستنباط الصفات قمنا بتطوير طريقة لإختيار الأفضل من هذه الصفات والأكثر قدره على التمييز بين الخبيث والحميد والسليم من أنواع ما يحدده الأطباء من أماكن مشبوهه في الماموجرام. طريقة الإختيار تعتمد على إستخدام طريقتي البحث التسلسلي والبحث العشوائي.

تم تطبيق الفكرة وتعليم الحاسوب باستخدام عدد من صور الأشعة بلغ عددها 410 صورة تم الحصول عليهم من قواعد البيانات الرقمية للماموجرام بالأشعة السينية من جامعة فلوريدا. وكانت النتيجة إختيار 31 صفه للأورام المحدده، منها 18 صفة إستنبطت بطريقتنا الجديدة المبتكره. وتم تقييم النتائج بإستخدام منحني روك المشهور وجدول التعارضات بين تصنيف طريقتنا المقترحة للصور الماموجراميه وبين حقيقة التصنيف. وكان الناتج بأن 95.1% من حالات الأورام الخبيثه، 91.7% من الأورام الحميده و 100% من الأنسجه السليمه تم تصنيفها بشكل صحيح. ونسبة الدقه وصلت إلى 96.3%.

Mammogram Computer-Aided Diagnosis

Belal Kamal Elfarra

ABSTRACT

Computer-aided diagnosis (CADx) is used to help radiologists in interpretation mammograms and is usually used as a second opinion by the radiologists. Improving CADx increases the treatment options and a cure is more likely. The main objective of this research is to enhance and introduce a new method for feature extraction and selection in order to build a CADx model to discriminate between cancers, benign, and healthy parenchyma. For feature extraction, we use both human features, which are obtained by Digital Database for Screening Mammography (DDSM), and computational features. For computational feature extraction, we enhance and use two pre-existed feature extraction methods, which are the Run Difference Method (RDM) and the Spatial Gray Level Dependence Method (SGLDM), and we propose a new feature extraction method called Square Centroid Lines Gray Level Distribution Method (SCLGM). Then, we evaluate and introduce a new method for feature selection by running both of forward sequential and genetic algorithm search methods individually. Later we evaluate the results.

Experimental results are obtained from a data set of 410 images taken from DDSM for different types.

Our method select 31 features from 145 extracted features; 18 of the selected features are from our proposed feature extraction method (SCLGM). We used both Receiver Operating Characteristics (ROC) and confusing matrix to measure the performance.

In training stage, our proposed method achieved an overall classification accuracy of 96.3%, with 92.9% sensitivity and 94.3% specificity. In testing stage, our proposed method achieved an overall classification accuracy of 89%, with 88.6% sensitivity and 83.3% specificity.

Key words: Breast Cancer, Mammogram, Feature Extraction, Feature Selection, Computer Aided Diagnosis, Genetic Algorithm, Forward Sequential.

Chapter 1

Introduction

1.1 Background

Breast cancer is the most common form of cancer among women and is the second leading of death after lung cancer. The American Cancer Society [1] estimates that in 2011 approximately 230,480 women in the US will be diagnosed with tumor breast cancer, and about 39,520 women will die from breast cancer. In Gaza, the breast cancer is the most common malignancy, accounting for more than 30% of all cancer in females. The incidence rate among the women in Gaza is 15.6 per 100,000 persons [2]. Early detection and diagnosis of breast cancer increase the treatment options and a cure is more likely. One of the most effective tools for early detection of breast cancer is the mammography, which is a screening tool used to examine the human breast by using low-dose amplitude X-rays.

Many countries encourage routine mammography of older women as a screening method to diagnose early breast cancer. Two views of each breast are recorded; the CranioCaudal (CC) view –head to foot view-, and a MedioLateral Oblique (MLO) view –angled side view. Figure 1-1 shows two views taken for case “A_1004 Right”: CranioCaudal view (left) and mediolateral oblique view (right).

The radiologists use these two views to search for any abnormalities; there are three signs usually radiologists look for:

- Masses, which are defined as space-occupying lesion seen in at least two different projections [3]. Masses are described by their shape and margin characteristics (Figure 1-2).
- Clusters of microcalcifications, which are very small bits of calcium, appearing on a mammogram as small bright spots. These spots are distributed in group clusters consisting of three or more spots in an area of about 1 cm² (Figure 1-3).

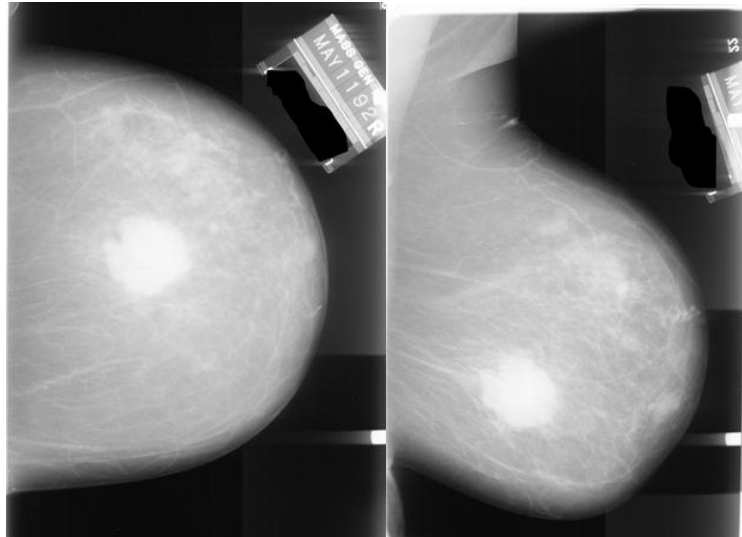


Figure 1-1: In mammogram two views are taken for each breast.

- Architectural distortions, which are defined as follows: “The normal architecture is distorted with no definite mass visible. This includes speculations radiating from a point, and focal retraction or distortion of the edge of the parenchyma [3]” (Figure 1-4).

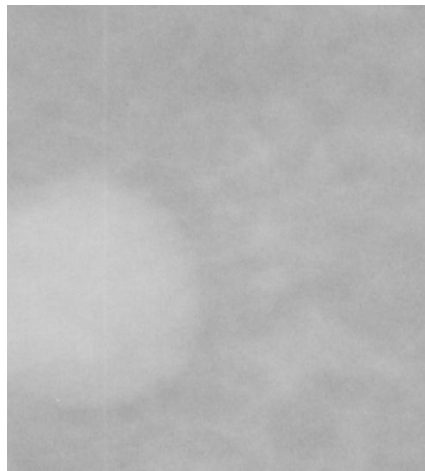


Figure 1-2: An example of a rounded circumscribed mass, this image is the case “A_0364 left” obtained from Digital Database for Screening Mammography DDSM.

A mammogram that has one sign of these abnormalities is not necessarily classified as malignant. Shapes, margins, size, and other features of the suspicious region have to be analyzed with experienced radiologist to classify the detected suspicious region into benign or malignant categories.

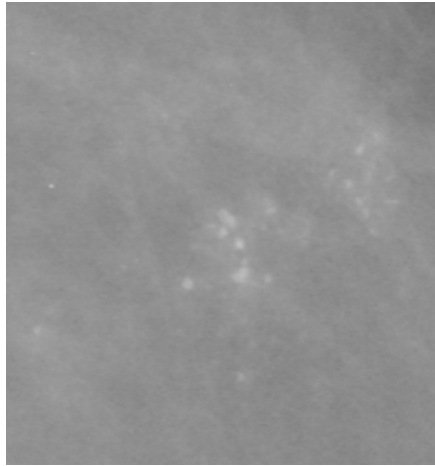


Figure 1-3: An example of microcalcification shown as small bright spots, this image is the case “A_1115 Right” obtained from Digital Database for Screening Mammography DDSM.

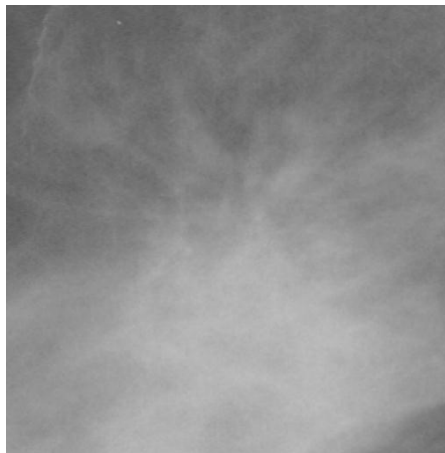


Figure 1-4: An example of a mass with Architectural distortions shape, this image is the case “A_1478 left” obtained from Digital Database for Screening Mammography DDSM.

Figure 1-5 and Figure 1-6 shows a schematic diagram of some mass shapes and boundary characteristics. The probability of malignancy with speculated and indistinct boundaries are greater than circumscribed masses.

Visual interpretation of a mammogram is a tedious and fatiguing process that generally requires a magnifying glass. The abnormality may be overlooked in a way that for each thousand cases we have only three to four cancerous. So, the probability of false negatives is high. Here the radiologists fail to detect 10% to 30% of cancers. Two thirds of these false negative results are due to missed lesions that are evident retrospectively [4].

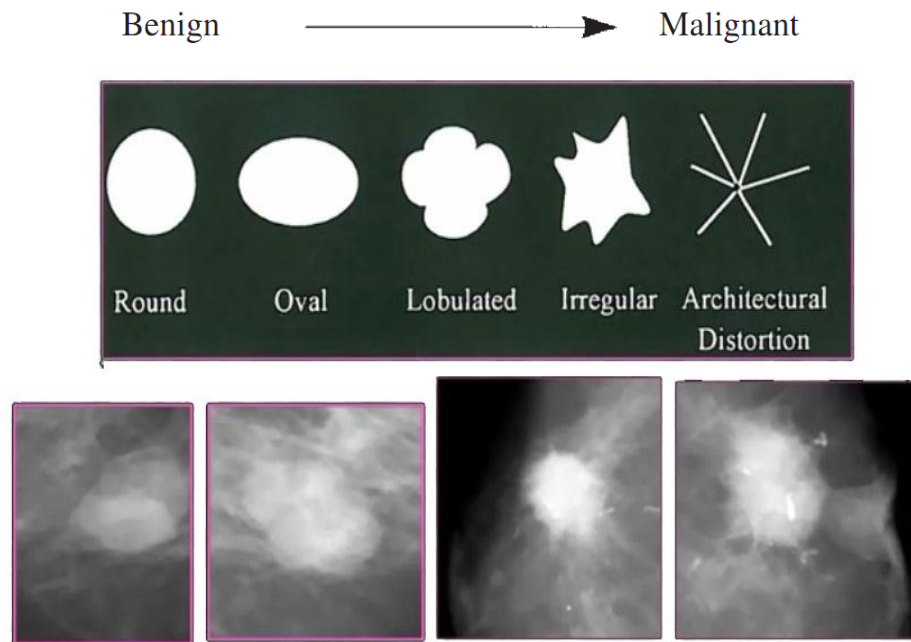


Figure 1-5: Some of mass shapes.

Also, a significant level of false positives were reported, that the positive predictive value is less than 35% which means a high proportion of biopsies are performed on benign lesions. Avoiding benign biopsies would spare women anxiety, discomfort, and expense.

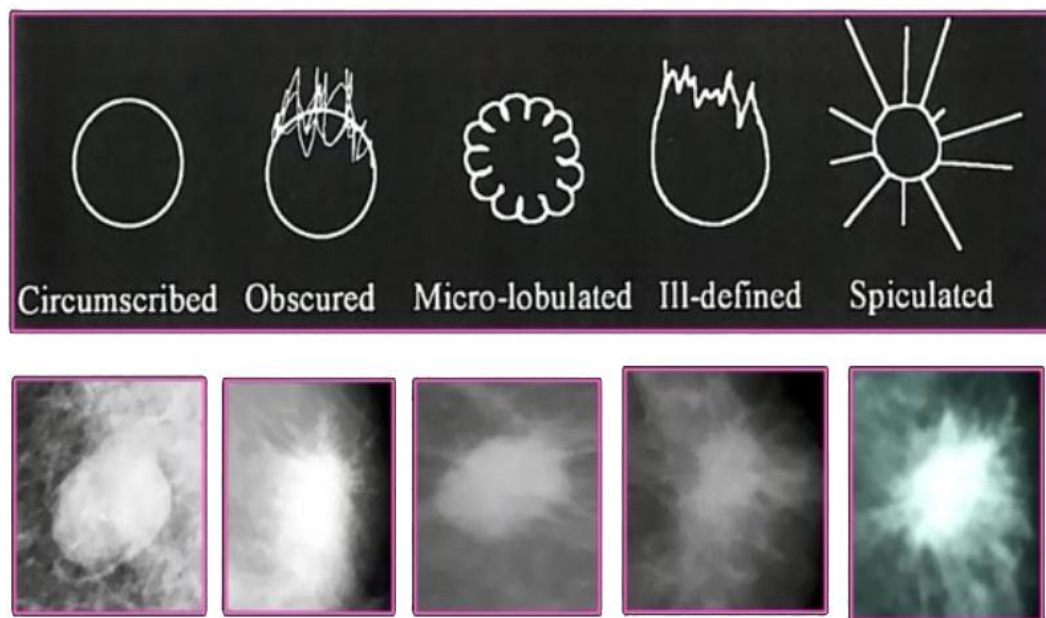


Figure 1-6: Mass margins.

Two systems have been developed to help the radiologists in reading mammogram. The first system is computer-aided detection (CAdE) which is used as a second reader

and the final decision is made by the radiologist. In recent studies, CAdE has improved radiologists' accuracy of detection of breast cancer [5, 6].

The second system is computer-aided diagnosis (CAdx) which classifies the detected regions into malignant or benign categories to help the radiologists in recognizing the next step, biopsy or short-term follow-up mammography.

Most detection algorithms consist of two stages. In the first stage, the aim is to detect suspicious lesions at a high sensitivity. In the second stage, the aim is to reduce the number of false positives without decreasing the sensitivity drastically.

Most diagnosis algorithms of CAdx consist of one stage. It begins with a lesion region or a region of interest (ROI) that contains the abnormality and outputs the likelihood of malignancy or a management recommendation.

For the two systems, each stage consists of five steps: preprocessing, segmentation, feature extraction, feature selection, and classification. These steps are explained in detail in Chapter3.

1.2 Research Overview

In this section, we present detailed information about this thesis. We start by identifying the importance of the optimization of the feature selection for mammogram diagnosis including the motivations behind our study, objectives to be accomplished, methodology that has been followed, our contributions throughout this work, and finally, we show the content of this research.

1.2.1 Motivation

As mentioned before, early detection and diagnosis of breast cancer increase the treatment options and a cure is more likely. Mammogram is an active tool used for early detection but 10%-30% of women who have the disease and undergo mammography have negative mammograms. Two thirds of these false negative cases were evident retrospectively [4]. These mistakes in the visual interpretation are due to poor image quality, eye fatigue of the radiologist, subtle nature of the findings, or lack experienced radiologists especially in third-world regions. Nowadays the computer-aid plays the main role in early detection and diagnosis of breast cancer. Increasing confidence in the diagnosis based on mammograms would, in turn, decrease the

number of patients with suspected breast cancer who have to undergo surgical breast biopsy, with its associated complications.

1.2.2 Objectives

The performances of different classifying algorithms can come to the same level in detecting mammogram lesions with the same features from the same mammographic database [7]. The performance of CADx depends more on the optimization of the feature selection than the classification methods. However, the feature space is very large and complex due to the wide diversity of the normal tissues and the variety of the abnormalities. Using excessive features may degrade the performance of the algorithm and increase the complexity of the classifier. For this, the main goal of this research is to evaluate methods for feature extraction, introduce other feature extraction techniques and enhance feature selection method to have best feature that guarantee the enhancement of classification with less dimension.

1.2.3 Methodology

In this research we evaluate methods for feature extraction, selection, and classification in order to build a CADx model to discriminate between some indicative patterns associated with cancers, benign, and healthy parenchyma. To grasp that, this research was conducted in the following phases:

- Preprocessing digital images of mammogram.
- Manual segmentation to the region of interest ROI which marked by experienced radiologist.
- Automated segmentation of the mass objects from the manual segmented ROI.
- Extracting features including: human extracting, computer extracting (texture descriptors) using pre-exist methods and introducing new feature extraction techniques.
- Selecting best features using sequential and randomized search strategies.
- Implementing the classification using Feed-Forward Artificial Neural Network with back-propagation (FF-ANN) algorithm.

1.2.4 Contributions

The contributions of this research are highlighted hereunder:

- Enhancing two pre-existing feature extraction technique, which are the Run Difference Method (RDM) and the Spatial Gray Level Dependence Method (SGLDM). These two methods extracted 60 features, 9 of these 60 features are selected as a significant features.
- Proposing a new feature extraction method called Square Centroid Lines Gray Level Distribution Method (SCLGM). By using this technique we extract 79 features, 18 of them selected as a significant feature.
- Enhancing and introducing a new feature selection method based on two types of search methods: forward sequential method and genetic algorithm method. These methods run individually and the total selected features are 31 features shared by 13 features.
- Decreasing the false positive prediction (FPP).
- Decreasing the percent of false negative prediction.
- As a result, 95.1% of cancer cases, 91.7 of benign cases and 100% of normal cases are correctly classified. The total accuracy is 96.3%.

1.2.5 Organization of This Thesis

The rest of this thesis is organized as follow: in Chapter two, we talk about some of the related work in the topic of our thesis. In Chapter three, we introduce an overview about computer aided diagnosis system; we talk in brief about CADx steps and performance metrics. The enhancement of pre-existing feature extraction techniques and our proposed feature extraction technique are described in detail in Chapter four. In Chapter five we preview the experiments and results, and then the conclusion and the suggested future work are viewed in Chapter six.

Chapter 2

Related Work

2.1 Segmentation

In segmentation step, the suspected objects are separated from background, this step is extremely important as the success of the classification algorithm depends on this stage [8]. The two major categories of segmentation methods are region growing and discrete contour models.

In [9], histogram based thresholding is used in segmentation step, where pixels with gray level values greater than 40 are retained while all others are set to zero, then statistical features are extracted from this region which signify the important texture features of breast tissue. These features are fed to the support vector machine (SVM) classifier to classify its tissue. By using threshold for segmentation in this method, some important information may be lost and other normal region will be added to the ROI such as nipple and lactation ducts. This addition will have bad effect on the results.

A supervised method for the segmentation of masses in mammographic images using texture features was proposed in [10, 11], which presented a homogeneous behavior inside the selected region. The results of this method are very good in segmentation but it requires a very large data to represent all types of cancers with all its possible shapes and margins.

Edge-detection, which is a traditional method for image segmentation, is used in [12] [13]. One of the main disadvantages of this method is its dependence on the size of objects and sensitivity to noise, Further, since conventional boundary finding relies on changes in the grey level, rather than their actual values, it is less sensitive to changes in the grey scale distributions over images as against region based segmentation.

Reddy [14] provide a novel method for image segmentation by using Gaussian low pass filter. They succeed in removing the boundary errors “false class” using combination of NN classifier and N-ary morphological operator. But, in this method

the difficulties of the classifier's designing increased as the number of classes in the image is increasing.

In [15], a new computerized method with a number of unique characteristics was developed to detect spiculated mass regions, and a simple spiculation index was applied to quantify mass spiculation levels. An active-contour model is used to refine the initial outlines, then the spiculated lines connected to the mass boundary are detected and identified using a special line detector. The computed area under the ROC curve, when applying the method to the data set, was 72%. Although this quantitative index can be used to distinguish between spiculated and circumscribed masses, the results also suggest that the automated detection of mass spiculation levels remains a technical challenge.

In [16], the lesion was segmented from the surrounding background using an adaptive region growing technique. Ninety-seven percent of the lesions were segmented using this approach.

In our experiments we use Region Growing technique; it is so suitable for mass segmentation [8].

2.2 Feature Extraction and Selection

The density tissues, which surround and mask the suspicious region, makes segmentation step very difficult. To overcome this problem we have to enhance feature extraction technique in order to obtain a large space of good features from which we could select discriminative features to improve the quality of lesion classification.

One of the feature extraction techniques that are mostly used in literatures is Gray-Level Co-occurrence Matrix (GLCM). In [17, 18] authors used five co-occurrence matrices statistics extracted from four spatial orientations, with $\theta \in (0, \pi/4, \pi/2, 3\pi/4)$ and pixel distance ($d=1$) to detect masses in mammograms. The extracted features won't be able to be discriminative with cases of cancers due to their non-uniform shape and margins. So our contribution is to increase the number of spatial orientations and increase the range of pixel distances.

Blot [19] compares the GLCM of the local ROI and compare it with the 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. This method will have bad results if ROI is surrounded by normal parenchyma with high gray level value such as nipple and lactation ducts.

A new approach of texture classification of 3-D Ultrasound (US) breast diagnosis using run difference matrix (RDM) with neural networks is developed in [20]. The accuracy reaches 91.9%. The results of this method are good, but not highly guaranteed because the used dataset is small; it includes just 54 malignant and 161 benign tumors.

Generally, the greater the number of features is, the higher the recognition rate will be. However, when the number of features is large but the number of training samples is small, features that have little or no discriminative information weaken the performance of classifiers. This situation is typically called the **curse of dimensionality** [21], in this situation we have to choose a feature subset yielding the highest performance.

It is very difficult to predict which feature or feature combinations will achieve better in classification rate. We will have different performances as a result of different feature combinations. In addition, using excessive features may degrade the performance of the algorithm and increase the complexity of the classifier. Relatively few features used in a classifier can keep the classification performance robust [22]. Therefore, we have to select an optimized subset of features from a large number of available features.

Two major methods for feature selection have been employed for CADx in mammography [23]: Stepwise feature selection and Genetic algorithm (GA). Stepwise feature selection are used in [24, 25, 26].

In [27], a feature selection method using stepwise Analysis Of Variance (ANOVA) Discriminant Analysis (DA) is used for classifying mammogram masses. This approach combines the 17 shape and margin properties of the mass regions and classifies the masses as benign or malignant. In ANOVA DA the discriminating

power of each feature is estimated based on grouping class variable. Principal component analysis (PCA) does feature extraction but it doesn't consider the grouping class variable. The experiment is performed on 300 DDSM database mammograms.

The main idea is to start with empty selected feature pool, and then at each step followed, one available feature is added to or removed from the selected feature pool with respect to the result of analyzing its effect on a selection criterion.

Wei [28] provide five selection criteria: (1) the minimization of Wilks' lambda, (2) the minimization of unexplained variance, (3) the maximization of the between-class F statistic value, (4) the maximization of Mahalanobis distance, and (5) the maximization of Lawley-Hotelling trace.

Most studies in mass detection [25, 26, 29] employed the minimization of Wilks' lambda as the selection criterion, which is defined as the ratio of within-group sum of squares to the total sum of squares.

In [26], the authors test all available selection criteria. A set of 340 features is reduced to 41 features with the stepwise feature selection. Stepwise feature selection techniques based on sequential search method and the main disadvantages of these techniques is it's availability of falling in local minima.

Veldkamp [30] used cluster shape features, cluster position features, and distribution features for the classification of calcifications. They used a sequential forward selection procedure for feature selection.

Genetic algorithm (GA) has been studied in [24, 26, 7, 29]. By GA we create a population of solutions based on the chromosomes and evolve the solutions by applying genetic operators such as mutation and crossover to find best solution(s) based on the predefined fitness function. The GA method with different fitness functions can reduce a set of 340 features to 39–62 features [15].

Tsujii [31] used the Karhunen-Loeve (KL) transform to reduce the dimensionality of the feature set. KL is not ideal and other techniques such as multiple discriminant analysis may be more appropriate for this task.

In [32], the authors developed a classification method that used only morphologic features. These features were designed after an in-depth study of the clinical characteristics of calcifications and produced very impressive results.

Aoki [21] assume that each feature subset has different discriminative information for different classes; they proposed a class decision tree classifier using class dependent feature subsets.

In [33], in this paper, authors classify mammogram images into normal image, benign image and malignant image. They used GLCM techniques to extract 26 features, and by a hybrid approach of feature selection they selected just 8 features, which are used by the decisions tree for mammogram classification. They applied their algorithms on just 113 mammogram image. Results are satisfactory but not highly guaranteed because the used dataset is so small and not enough. Another disadvantage in their work is about the extracted features which are so little, and this weakens the probability of selecting a discriminative features.

In our method we extract 145 features, and we use three datasets: the first dataset contains 410 mammogram images and is used for selecting features and creating CADx model and the others two datasets each of which contain 100 mammogram images and are used for testing the created model.

2.3 Classification

In [17] the classification is performed using each feature vector and linear discriminant analysis (LDA).

In [34], A new type of classifier combining an unsupervised and a supervised model was designed and applied to classification of malignant and benign masses on mammograms. The unsupervised model was based on an adaptive resonance theory (ART2) network which clustered the masses into a number of separate classes. The classes were divided into two types: one containing only malignant masses and the other containing a mix of malignant and benign masses. The masses from the malignant classes were classified by ART2. The masses from the mixed classes were input to a supervised linear discriminant classifier (LDA). In this way, some malignant masses were separated and classified by ART2 and the less distinguishable benign and malignant masses were classified by LDA.

For classification LDA does very good when the training data are well representative of the data in the system, and surely it is so difficult to have this type of data.

Decision tree is used as mammogram classifier in [21, 33]. The main advantage of using decision tree is its simplicity and less sensitivity to errors. But its results candidates to over-fit the training data, especially when the used training data is too small or have noise.

SVM classifier used in [9], two main advantages of using SVM, with SVM the solution is global and unique, and the geometric interpretation is simple. Also there are two main disadvantages with SVM, the first is that: SVM needs multiple binary SVMs to deal with multi-class, and the second is: SVMs are less prone to over-fitting, which means, SVM is not good with biggest problem.

Artificial Neural Network (ANN) used in [14, 20, 35], many advantages of using ANN, is well-suited to problems in which the training data corresponds to noisy and complex sensor data such as mammogram. It maintains non-linearity and it could deal with biggest problems. But it is suffering from multiple local minima; the problem of local minima could be solved by using techniques such as: stochastic gradient descent and k-fold.

A novel soft cluster neural network technique for the classification of suspicious areas in digital mammograms is presented in [36] in this method, and the generalization ability of the neural network is increased by providing a mechanism to more aptly depict the relationship between the input features and the subsequent classification as either a benign or malignant class. Soft clusters with least squares make the training process faster and avoid iterative processes which have many problems.

In our method we use Feed Forward Artificial Neural Network (FF-ANN) with back propagation technique.

Chapter 3

Background

3.1 CADx System

Computer-aided detection (CAD) consists of two stages: in the first stage it marks all abnormalities with high sensitivity, and then in the second stage it reduces these regions by eliminating the false positive regions. Radiologist use CAD system as a second reader for the mammogram, they compare the CAD's results with their first opinion, and then they mark these abnormalities as a region of interest (ROI), which is used as input to Computer Aided diagnosis system (CADx).

CADx is a very important tool, that it could aid radiologists in taking a decision about the next step: short-term follow-up or biopsy. Without CADx system, most decisions of inexperienced radiologist were in the side of biopsy to avoid false negative which lead to death, ignoring women stress, and biopsy costs.

As shown in Figure 3-1, CADx system consists of five steps :1- Preprocessing, 2- Segmentation, 3- Feature extraction, 4- Feature selection, 5- Classification. These steps will be described in detail in the following subsections.

3.1.1 Preprocessing

One of the reasons leading to difficulty of interpretation of mammogram is that: mass, calcification or any other types of cancer are hidden by the surrounded density tissues, which mask the suspicious region. Preprocessing step aims to enhance and to improve the quality of the original mammogram image and to make the following steps in CADx easier and more reliable.

Enhancement of mammogram can be done by increasing the contrast of mammogram which has been taken at low-dose X-ray, or by sharpening image edges and boundaries. Image enhancement includes gray-level and contrast manipulation, noise reduction, edge crisping and sharpening, filtering, interpolation and magnification, and so on.

The greatest difficulty in image enhancement is quantifying the evaluation criteria for enhancement. Image enhancement techniques can be improved if the enhancement criteria can be stated precisely. Often such criteria are application dependent [37].

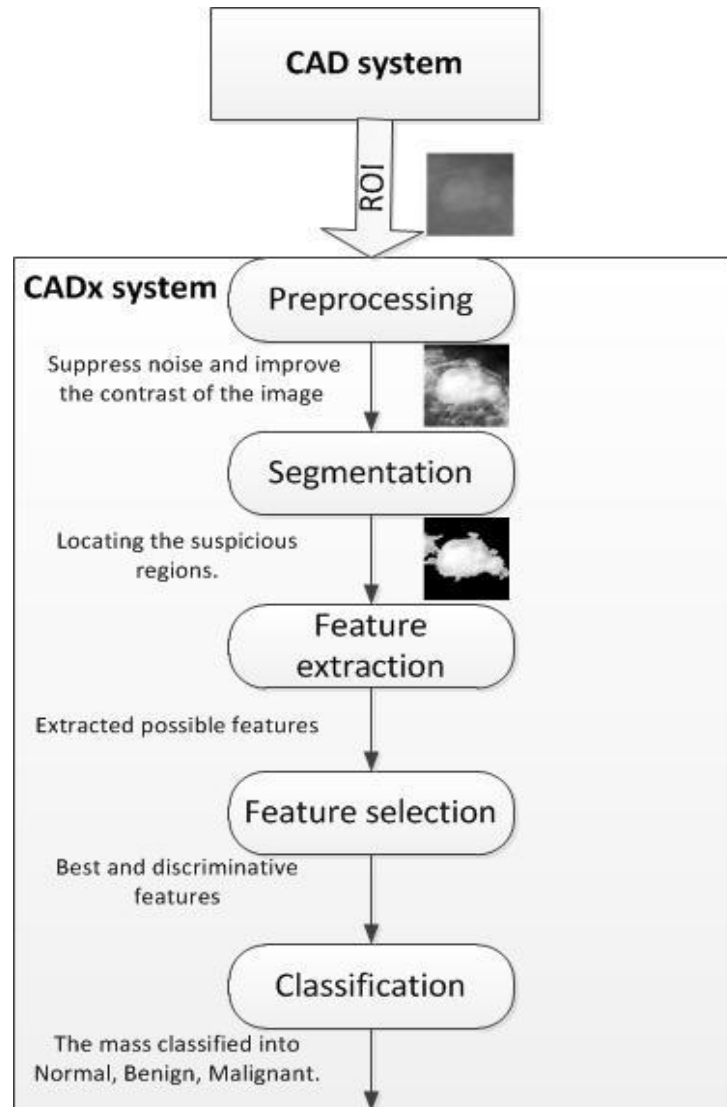


Figure 3-1: Mammogram Computer-Aided Diagnosis (CADx) steps.

Histogram equalization method is a simple global enhancement technique; it is used for the adjustment of the image contrast by using the image's histogram. By using this method, the intensities are distributed in a good way over the histogram, which allows the areas of lower local contrast to get better contrast.

This method is helpful and provides useful results with images at which the intensity of background and foreground are simple, both are bright or dark. In mammogram the abnormalities are often occluded or hidden in dense breast tissue, for this we prefer using Histogram equalization in preprocessing step.

3.1.2 Segmentation

To extract features from the abnormal regions we have to separate the suspicious regions from the background parenchyma, and this is the goal of segmentation step.

Segmentation of the mass can be done manually, semi-automated, or fully automated. In general, automated segmentation algorithms are based on one of the two basic intensity properties: discontinuity and similarity.

Algorithms based on discontinuity are suitable for calcification at which we have abrupt changes in intensity, but for mass types we use segmentation algorithms based on similarity. Region growing, and region splitting and merging are examples of similarity category segmentation [38].

Region splitting and merging

In this method, the image is subdivided into a set of arbitrary disjointed regions, and then split regions that have pixels, which are not within the criteria property –for example not having the same gray level- and merge any adjacent regions shared in their properties, this process continue until no more split or merge is allowed.

Region growing

Region growing, in general, is a procedure by which we group pixels or sub-regions into larger region based on predefined criteria [38]. We start with a set of “seed” points and from these points we start region growing by appending neighboring points that share these seed points in properties. Connectivity or adjacency must be taken in account.

In our research, we use region growing method, that after manually segmenting ROI which is marked by experienced radiologists, we identify a set of points at the center of each suspicious region and use these points as a seed from which we start region growing and stop when no more points are added to the region.

3.1.3 Feature Extraction

Features are defined as the characteristics of ROI, these features could be organized into morphologic features and texture features:

Morphologic features

Morphologic features are human notable and directly inspired by characteristics defined by radiologists. Shape, margin and density are examples of morphologic features.

Texture features

Texture features computed mathematically, which are not evident to human eye and not easily extracted visually. Gray level co-occurrence matrix (GLCM) is used to compute texture feature; it is a statistical method that considers the spatial relationship of pixels in the gray-level co-occurrence matrix. Each element (I, J) in the resultant GLCM is simply the sum of the number of times that the pixel with value I occurred in the specified spatial relationship to a pixel with value J in the input image [33]. The following are some examples of GLCM features: correlation, homogeneity, sum average, cluster prominence, maximum probability, variance, energy, contrast, entropy, and uniformity.

3.1.4 Feature Selection

The extracted feature space is very large and complex due to the wide diversity of the normal tissues and the variety of the abnormalities [39]. Using excessive features lead to what is called “curse of dimensionality” that degrades the performance of the classifier and increases the complexity. To improve the prediction accuracy and minimize the computation time, we have to remove the irrelevant, redundant and noisy features, and keep only a robust combination of features which will be a discriminative and helpful for a classifier. This procedure is called feature selection.

Feature selection step is very important, that the performance of diagnosis depends more on the optimized feature selection than the classification. A number of search procedures have been proposed for feature selection, which can be categorized in two categories: sequential feature selection and randomized feature selection.

Sequential feature selection

Many algorithms are classified as sequential search such as: Sequential Forward Selection, Sequential Backward Selection, Plus-1 Minus-r Selection, Bidirectional Search, and Sequential Floating Selection. The main idea of these algorithms is to add or remove features to the vector of selected features, and the iteration continues until all features are checked. Adding or removing a feature is based on one of many criteria functions that could be used. Misclassification, distance, correlation, and entropy are examples of criteria functions. The main disadvantage of these algorithms is that, they have a tendency to become trapped in local minima.

The pseudo code description of the Sequential Forward Selection method is given in Algorithm 3.1.

Algorithm 3.1: <i>Sequential Forward Selection</i>	
Purpose:	1. Selecting a K discriminative features from the set of extracted features (F)
Input:	2. The set of extracted features $F = \{f_1, f_2, f_3, \dots, f_n\}$, 3. The number of features to be select K , 4. The used criteria function $J()$.
Output:	5. the selected features set X_k
Procedure:	6. <u>begin</u> 7. <u>initialize</u> $x_1 = \operatorname{argmax}_{f_i \in F} J(f_i)$; $X = \{\}$; $k = 0$; 8. while $k < K$ do{ 9. $x_{k+1} = \operatorname{argmax}_{f_i \in (F - X_k)} J(f_i, X_k)$; 10. // $J(f_i, X_k)$ is the significance of a feature f_i in conjunction //with other features in the set X_k 11. $X_{k+1} = X_k \cup x_{k+1}$; 12. $k = k + 1$; 13. } 14. <u>end</u> ;

Randomized feature selection

These algorithms are based in randomness in their search procedure to escape falling in local minima. Random Generation plus Sequential Selection, Simulated Annealing

and Genetic Algorithms are some examples of algorithms based on randomness in their search.

In our experiments we use Genetic algorithm as a randomized feature selection. Figure 3-2 shows idea of the basic genetic algorithm. Each of the L classifiers in the population in generation k is represented by a string of bits of length N , called a chromosome (on the left). Each classifier is judged or scored according to its performance on a classification task, giving L scalar values f_i .

The chromosomes are then ranked according to these scores. The chromosomes are considered in descending order of score, and operated upon by the genetic operators of replication, crossover and mutation to form the next generation of chromosomes the offspring. The cycle repeats until a classifier exceeds the criterion score θ [40].

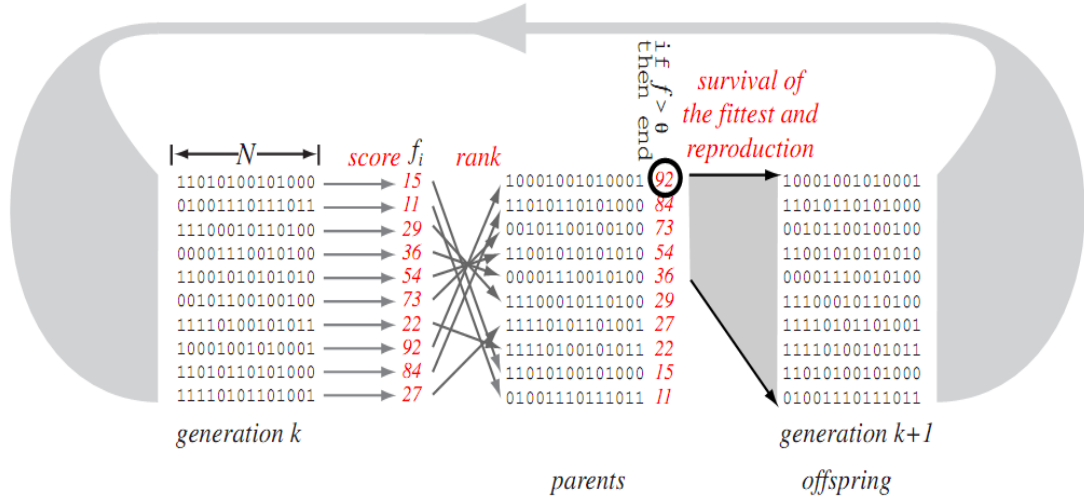


Figure 3-2: A basic genetic algorithm is a stochastic iterative search method [40]

The pseudo code description of the basic Genetic Algorithms method is given in Algorithm 3.2.

Algorithm 3.2: Basic Genetic algorithm [40]	
Purpose:	1. Finding the optimized solution for the problem using the same combination, recombination and mutation of approximated solutions.
Input:	2. Crossover probability P_{co} 3. Mutation probability P_{mut} 4. Population size: L -chromosomes- or classifier- by N -bits. 5. The used criteria function $Fit()$. 6. Fitness threshold θ .
Output:	7. The set of highest fitness chromosomes (best classifier)

Procedure:	8. <u>begin</u> 9. <u>do</u> 10. <i>Determine fitness of each chromosome, $Fit(i), i = 1, \dots, L$</i> 11. <i>Rank the chromosomes</i> 12. <u>do</u> 13. <i>Select two chromosomes with highest score</i> 14. <i>if ($Rand[0, 1) < P_{co}$) then</i> 15. <i>crossover the pair at a randomly chosen bit</i> 16. <i>else</i> 17. <i>change each bit with probability P_{mut}</i> 18. <i>Remove the parent chromosomes</i> 19. <i>until N offspring have been created</i> 20. <i>until Any chromosome's score $Fit()$ exceeds θ</i> 21. <u>return</u> <i>Highest fitness chromosome (best classifier)</i> 22. <u>end</u>
-------------------	---

3.1.5 Classification

Classification is the last step in CADx system; in this step the classifier uses the selected robust significant features obtained from the feature selection step as input to classify the abnormal lesion into normal, benign, or malignant. In the following, there are some examples of classifiers:

Support vector machine (SVM)

SVM is an example of supervised learning methods used for classification and regression analysis. SVM is a binary linear classifier that for each given input data it predicts which of two possible classes comprises the input. For example, in Figure 3-3 we have a set of training examples categorized into one of two categories, the SVM goal is to build a model that categorizes any new examples into one of the two categories. In the figure we have three hyperplane: H3 (green) does not separate the two classes, but H1 and H2 do.

Our goal is to search for the direction that gives the maximum possible margin. Margin which is defined as the hyperplane leaves from both classes [4]. The margin of H_2 is the maximum margin, so for H_2 , the best hyperplane is used to separate the two classes. After locating the best hyperplane with maximum margin, SVM trained it with samples from two classes.

Figure 3-4 shows some samples located on the margin, these samples are called the support vector. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

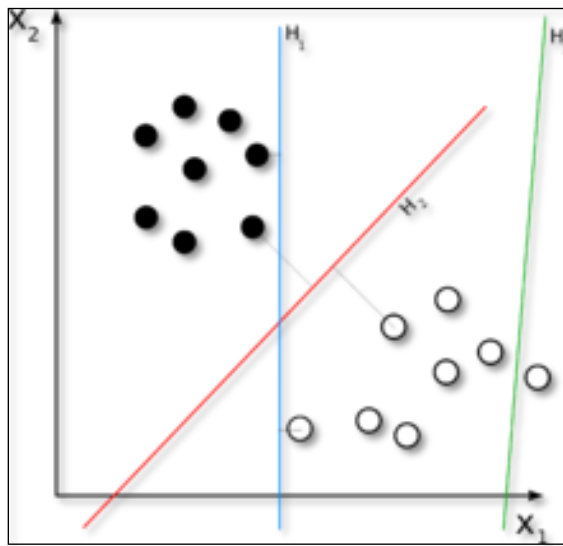


Figure 3-3: SVM will choose the hyperpalne H2 by which we have best margin [41].

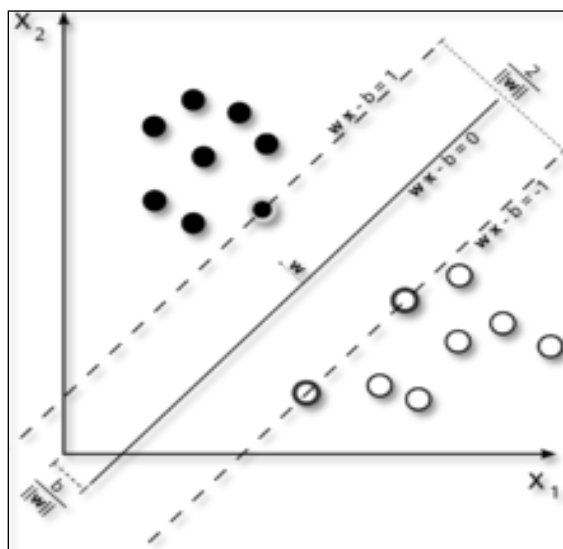


Figure 3-4: Samples that located on the margin called the support vector [41].

Artificial neural network (ANN)

ANN is an appropriate simulation of the nervous system. It consists of three or more interconnected layers; each layer consists of one or more nodes, which are called neurons, or nerve cells. Each neuron receives a number of inputs comes via a connection -presented as arrows in Figure 3-5.

The arrows indicate the direction of information flow, feeding information forward from input to output, each arrow has a strength or (weight). Each neuron also has a single threshold value (w_0), this threshold subtracted from the sum of incoming weights to compose the activation of the neuron. The activation signal is passed through an activation function (transfer function) to produce the output of the neuron y , y could be represented by the following function:

$$y = f(x, w) \quad 3.1$$

Where x is the inputs x and w is the connections weights; in the following we view two examples of neurons functions:

Sigmoidal neuron:

$$y = \frac{1}{1 + e^{-w^t x - a}} \quad 3.2$$

Gaussian neuron:

$$y = e^{-\frac{\|x-w\|^2}{2a^2}} \quad 3.3$$

Figure 3-6 shows an example of three-layer of ANN, one layer for each: inputs, hidden and output nodes. Note all nodes need to be connected to every node in the next layer.

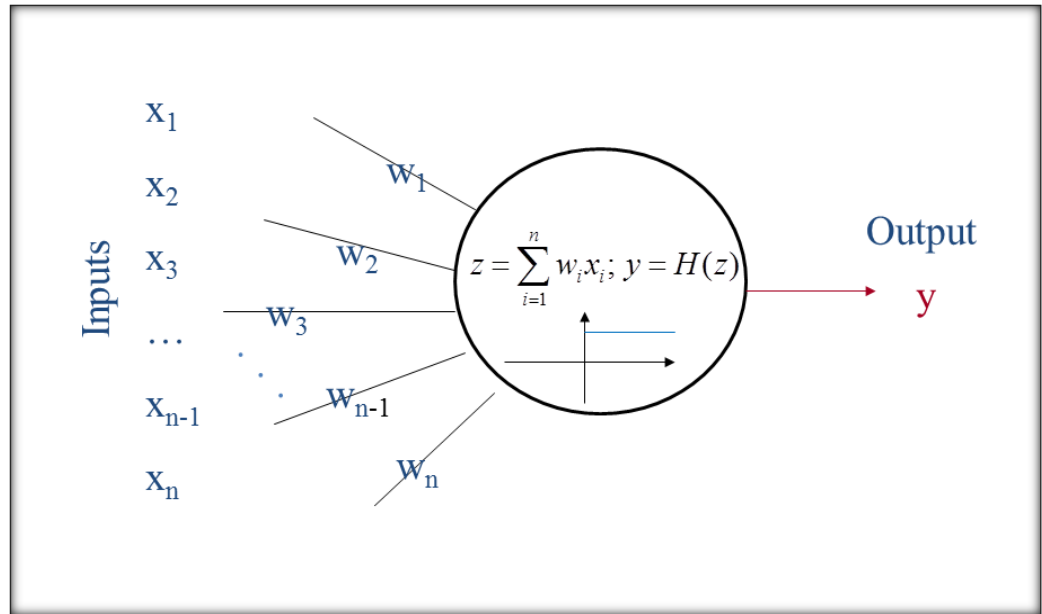


Figure 3-5: the structure of one neurons in ANN

Prediction with ANNs involves two steps, one is training and the other is learning. Training of Feed forward artificial neural networks (FFANN) is normally performed in a supervised manner. The success of training is greatly affected by proper selection of inputs. In the learning process, a neural network constructs an input-output mapping, adjusting the weights and biases at each iteration based on the minimization or optimization of some error measured between the output produced and the desired output. This process is repeated until an acceptable criterion for convergence is reached. In the back propagation (BP) algorithm, the input is passed layer through layer until the final output is calculated, and it is compared to the real output to find the error. The error is then propagated back to the input adjusting the weights and biases in each layer. In the standard BP learning algorithm the sum of square errors is minimized.

3.2 Performance Metrics

We measure the accuracy using the following equation:

$$\text{Accuracy} = \frac{(\text{\#of test observations at which predicted value} = \text{target})}{\text{total test observations}} \quad 3.4$$

The accuracy equation is not enough as a metric of performance, so we use the Receiver Operating Characteristics (ROC) curve which is the most useful and clear possibility for organizing classifiers and visualizing their quality [42, 8].

Figure 3-7 shows a two dimensional graphical plot, by which the y-axis represents the sensitivity, or true positive rate, and the x-axis represents the false positive rate (1-specificity). Sensitivity and specificity are computed by the following:

$$Sensitivity = \frac{(\#of\ true\ positive\ classification)}{(\#of\ cancerous\ lesions)} \quad 3.5$$

$$Specificity = \frac{(\#of\ true\ negative\ class)}{(\#of\ benign\ lesions)} \quad 3.6$$

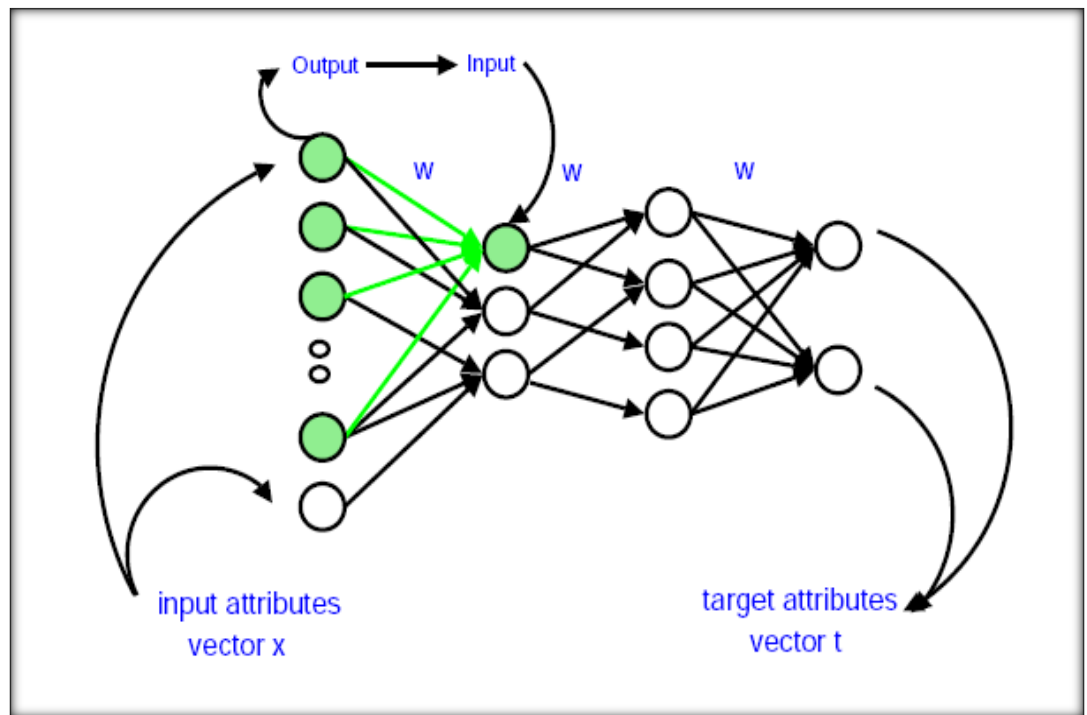


Figure 3-6: An example of three-layer of ANN [41].

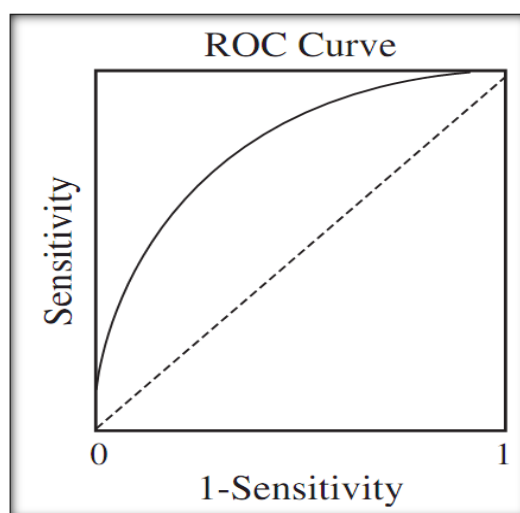


Figure 3-7: ROC curve.

The performance of the classifier could be measured with respect to the area under the ROC curve, that we have better performance as curve area increased. Beside ROC curve, we use also the confusion matrix, which represents the summation of true and false of both positive and negative instances as demonstrated in Table 3-1.

Table 3-1: The structure of Confusion matrix.

	positive (Predicted)	negative (Predicted)	Accuracy
positive (actual)	Summation of true positive	Summation of false negative.	
negative (actual)	Summation of false positive	Summation of true negative.	
Accuracy			Total accuracy

Chapter 4

Proposed Feature Extraction and Selection Techniques

4.1 Introduction

When radiologists need to diagnose a mass in mammogram, they look for some significant features that discriminate malignant from benign masses. These visual features -which are based on shape, size and margin - could have different interpretation based on radiologist's opinion and experience. To solve the problem of these different interpretations, more discriminative features should be extracted. Computer provides multiple methods for obtaining these discriminative features, which are done in two steps:

Feature extraction: this step is responsible for extracting all possible features that are expected to be effective in diagnosing a ROI in mammogram, without concerning the disadvantages of excessive dimensionality.

Feature selection: as mentioned before this step is responsible for reducing the dimensionality by removing redundant features and searching for the best significant features to avoid the curse of dimensionality.

In this Chapter, we will introduce our proposed methods for feature extraction and selection.

4.2 Feature Extraction

In this section, we describe the computational and the human features extractions. With computational feature extraction, we use and modify two of known methods, which are Spatial Gray Level Dependence Method (SGLDM) and Run Difference Method (RDM) to enhance their power in describing textural characteristics of the mass patterns. In addition to these two methods, we propose a new feature extraction method; we call it "Square Centroid Lines Gray Level Distribution Method (SCLGM)".

More details about these three methods and human feature extraction will be introduced in the following subsections.

4.2.1 Spatial Gray Level Dependence Method (SGLDM)

Also called Co-occurrence matrix, it is a statistical method proposed by Julesz in 1975, who induced that: “no texture pair can be discriminated if they agree in their second-order statistics” [43]. Then Julesz used the definition of the joint probability distributions of pairs of pixels for texture feature extraction.

The SGLDM matrix is formed by computing the number of occurrences of each pixel with gray level i that are away by distance d from any pixel with gray level j in a direction defined by angle θ . The choice of distance and angle combination, as well as the quantization level, is somewhat arbitrary. Figure 4-1 shows the co-occurrence for one pixel (yellow pixel) with $d=3$ pixel and $\theta \in \{0, \pi/4, 2\pi/4, 3\pi/4\}$.

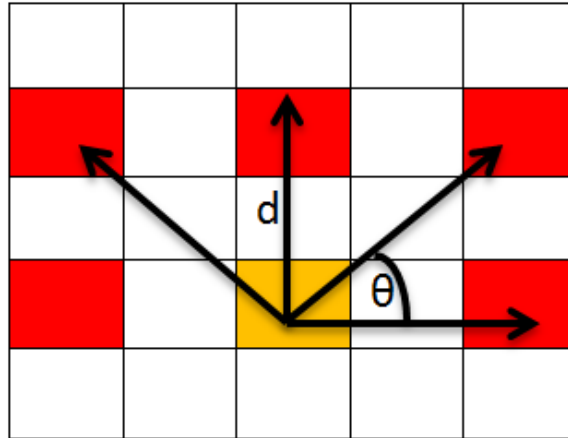


Figure 4-1: The co-occurrence for one pixel (yellow pixel) with $d=3$ pixel and $\theta \in \{0, \pi/4, 2\pi/4, 3\pi/4\}$.

Figure 4-2 demonstrates how to compute co-occurrence matrix for an image with dimension 3×3 , at $d=1$ and $\theta \in \{0^\circ, 45^\circ\}$. As shown in the example, the largest value of the used gray level in the image (left) is four, so the dimension of co-occurrence matrix will be 5×5 . Now for each pair (i, j) of gray levels in the co-occurrence matrix we tabulate the number of occurrences of this pair in the image with distance d and angle θ .

The choice of distance and angle combination, as well as the quantization level, is somewhat arbitrary [44]. Usually, the selection of parameter d is based on texture granularity, which means that with fine texture we use small values for d , and with

coarse texture –as in malignant mass- large values of d will be used. In our method, we compute the co-occurrence matrix for a square ROI at eight different angles: $\theta \in \{0^\circ, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, \text{ and } 7\pi/8\}$, and at all distances $d \in \{1, 2, 3, \dots, L/2\}$, where L is the length of ROI's side. We use these variant values of distances and angles to have discriminative features for the mass pattern to distinguish between benign or normal mass and malignant mass.

As a result, eight co-occurrence matrices are generated for each value of distance d ; these matrices are summed to obtain a rotation invariant matrix M^d :

$$M_{i,j}^d = \sum_{\theta} M_{i,j}^{\theta,d} \quad 4.1$$

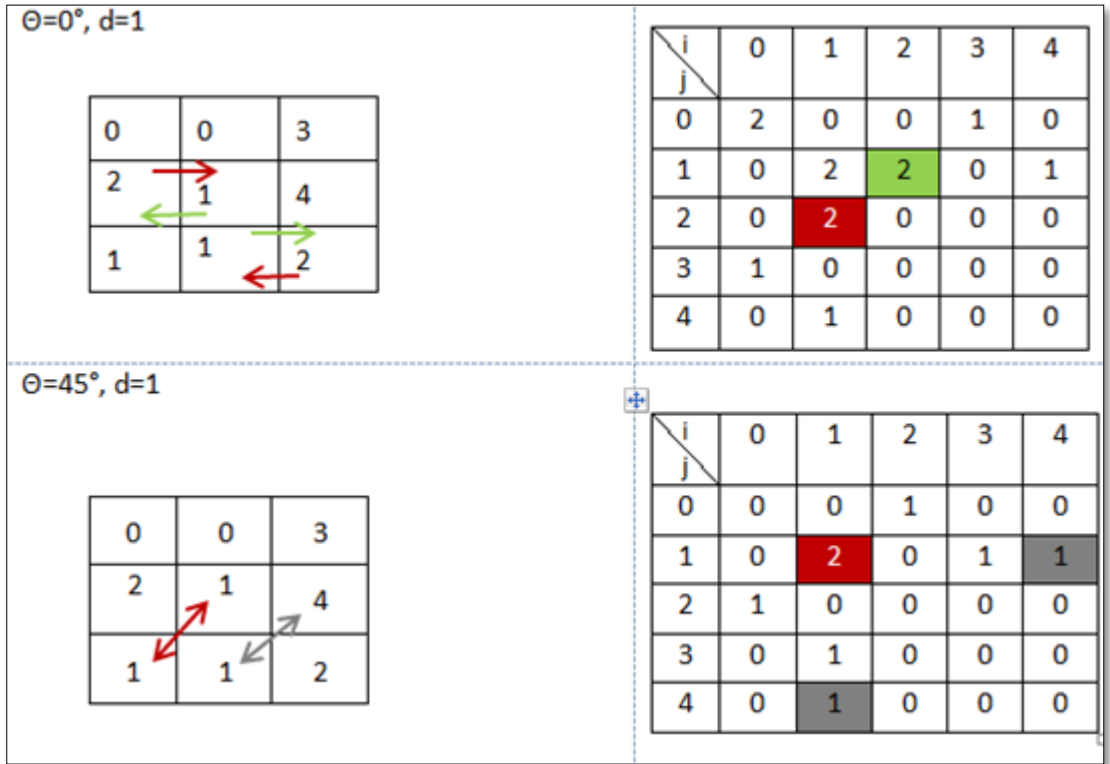


Figure 4-2: Two example of co-occurrence matrix with $d=1$ and $\theta=0^\circ$ (upper) and $\theta=45^\circ$ (lower).

Then we extract the eight of 14 co-occurrence features proposed by Haralick [45]. These features are computed by equations that follow, where G denotes the number of gray levels, σ_i^d, σ_j^d and μ_i^d, μ_j^d denotes the standard deviations and the mean of the sum of co-occurrence matrix at distance value d .

1- **Correlation:**

$$x_1 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \left[\frac{(i - \mu_i^d)(j - \mu_j^d)}{\sigma_i^d \sigma_j^d} \right] \quad 4.2$$

2- **Homogeneity:**

$$x_2 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \left[\frac{M_{i,j}^d}{1 + (i - j)^2} \right] \quad 4.3$$

3- **Variance:**

$$x_3 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu^d)^2 \quad 4.4$$

4- **Contrast:**

$$x_4 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - j)^2 \quad 4.5$$

5- **Energy:**

$$x_5 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (M_{i,j}^d)^2 \quad 4.6$$

6- **Maximum probability:**

$$x_6 = \max_{i,j} M_{i,j}^d \quad 4.7$$

7- **Sum average:**

$$x_7 = \frac{1}{2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i + j) \quad 4.8$$

8- **Cluster prominence:**

$$x_8 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu_i^d + j - \mu_j^d)^4 \quad 4.9$$

If we apply these eight measures for each value of $d \in [1, L/2]$, we have $(8 \cdot L/2)$ features. This is very large data, most of which are redundant and not useful. So we

use the following functions for each measure - co-occurrence feature – over all values of d , these functions are:

1- **Mean:**

$$x_{n+8} = \frac{1}{L/2} \sum_{d=1}^{L/2} x_n^d, \quad \text{for all } n \in \{1,2,3,4,5,6,7,8\} \quad 4.10$$

2- **Mean absolute deviation:**

$$x_{n+16} = \frac{1}{L/2} \sum_{d=1}^{L/2} |x_n^d - x_{n+8}|, \quad \text{for all } n \in \{1,2,3,4,5,6,7,8\} \quad 4.11$$

3- **Minimum:**

$$x_{n+24} = \text{Min}_{d=1}^{L/2} x_n^d, \quad \text{for all } n \in \{1,2,3,4,5,6,7,8\} \quad 4.12$$

4- **Maximum:**

$$x_{n+32} = \text{Max}_{d=1}^{L/2} x_n^d, \quad \text{for all } n \in \{1,2,3,4,5,6,7,8\} \quad 4.13$$

5- **Variance:**

$$x_{n+40} = \frac{1}{L/2} \sum_{d=1}^{L/2} (x_n^d - x_{n+8})^2, \quad \text{for all } n \in \{1,2,3,4,5,6,7,8\} \quad 4.14$$

Where x_{n+8} is the mean of feature x_n^d over all $d \in \{1, 2, \dots, L/2\}$.

6- **Skewness:**

$$x_{n+48} = \frac{\frac{1}{L/2} \sum_{d=1}^{L/2} (x_n^d - x_{n+8})^3}{\left(\frac{1}{L/2} \sum_{d=1}^{L/2} (x_n^d - x_{n+8})^2 \right)^{\frac{3}{2}}}, \quad \text{for all } n \in \{1,2,3,4,5,6,7,8\} \quad 4.15$$

The pseudo code description of Spatial Gray Level Dependence Method is given in Algorithm 4.1.

In steps 7 – 10 of the algorithm, we create eight images, each of which represents the origin image of the mammogram rotated by $\theta \in \{0^\circ, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, \text{ and } 7\pi/8\}$. The purpose of these steps is to create the eight SGLDM matrices in one direction ($\theta=0$).

Algorithm 4.1: The pseudo code of <i>SGLDM</i> technique.	
Purpose:	1. Extracting 56 features from the mammogram images.
Input:	2. Enhanced segmented objects with black background (I). 3. The used direction for co-occurrence $\theta = \{0^\circ, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, 7\pi/8\}$ 4. Features To be Computed $FTC = \{ \text{correlation, homogeneity, sum average, cluster prominence, maximum probability, variance, energy, contrast} \}$
Output:	5. The <i>SGLDM</i> extracted features vector of the input image I.
Procedure:	6. <u>begin</u> 7. <u>for each angle in θ do</u> 8. create $I(\theta)$, which is the image rotated by angle θ . 9. $I(\theta) = \text{crop } I(\theta)$ to get the inscribed squares within the $I(\theta)$. 10. <u>end loop;</u> 11. $L = \text{image side};$ 12. <u>for all distances $d \in \{1, 2, 3, \dots, L/2\}$ do</u> 13. $Cooc = 0;$ //Cooc is the sum of co-occurrence matrices 14. <u>for each angle in θ do</u> 15. $Cooc = Cooc + \text{cooccurrence matrix of } I(\theta)$ 16. <u>end loop;</u> 17. <u>for $i = 1$ to $\text{length}(FTC)$ do</u> 18. $S_i(d) = \text{computeFeatures}(Cooc, FTC(i));$ 19. <u>end loop;</u> 20. <u>end loop;</u> 21. $n = \text{length}(FTC);$ //n is the number of measures to be computed //features statistics: 22. <u>for $i = 1$ to $\text{length}(FTC)$ do</u> 23. $S_{i+n} = \text{mean}(S_i);$ 24. $S_{i+2n} = \text{MAD}(S_i);$ 25. $S_{i+3n} = \text{minimum}(S_i);$ 26. $S_{i+4n} = \text{maximum}(S_i);$ 27. $S_{i+5n} = \text{variance}(S_i);$ 28. $S_{i+6n} = \text{skewness}(S_i);$ 29. <u>end loop;</u> 30. $features = [S];$ 31. <u>return features;</u> 32. <u>end</u>

In step 18, we extract the co-occurrence features from *Cooc* matrix which is the sum of the eight SGLDM matrices computed at step 15.

4.2.2 Run Difference Matrix (RDM)

RDM is based on the estimation of the probability density function of the gray level differences in an image. The run difference matrix comprised the gray-level difference along with a distance between the pixels, when the displacement vector between two pixels is given [20].

With respect to Figure 4-3, we can compute the displacement vector \bar{D} using the following equation:

$$\bar{D} = [\Delta x, \Delta y] = [x_a, y_b] - [x_n, y_m] \quad 4.16$$

By using Figure 4-3, RDM could be defined as a total number of pixels pair in ROI with distance r and gray level difference with the given direction θ :

$$RDM(r, g_{dif} | \theta) = \# \{ ((x_a, y_b), (x_n, y_m)) : ((x_a, y_b), (x_n, y_m)) \in ROI, |G(x_n, y_m) - G(x_a, y_b)| = g_{dif} \} / N \quad 4.17$$

Where $G(x, y)$ is the gray level value of the pixel (x, y) ,

$$r = \sqrt{\Delta x^2 + \Delta y^2}, \quad \theta = \tan^{-1} \left(\frac{\Delta y}{\Delta x} \right) \quad 4.18$$

And N is used for normalization and it is the total number of all pixels pairs in ROI:

$$N = \{ ((x_a, y_b), (x_n, y_m)) : ((x_a, y_b), (x_n, y_m)) \in ROI \} \quad 4.19$$

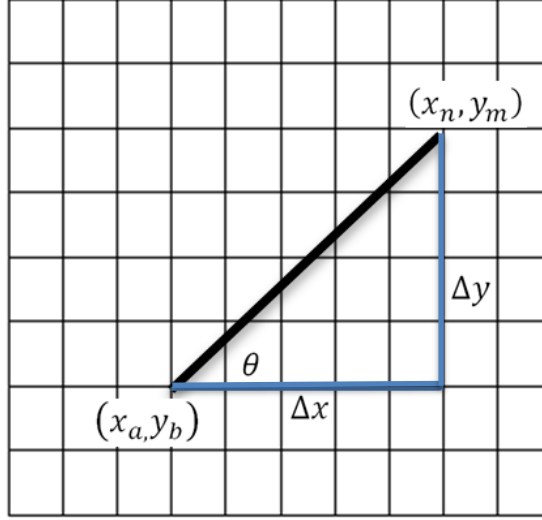


Figure 4-3: Run Difference Matrix (RDM) is a function of r and level difference with the given direction θ .

In cancer cases, the shape and margins are non-uniform. So we propose to extend RDM matrices for eight different angles $\theta \in \{0^\circ, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, \text{ and } 7\pi/8\}$. As a result, eight RDM matrices are generated; these matrices are summed to obtain a rotation invariant matrix M_{RDM} :

$$M_{RDM} = \sum_{\theta} RDM(r, g_{dif} | \theta) \quad 4.20$$

The obtained matrix M_{RDM} is not designed for the extraction of features, but we can use the three characteristic vectors defined along with the original run difference matrix to extract features. These characteristics are the DGD (distribution of gray level difference), the DOD vector (distribution of average difference), and the DAD (distribution of average distance) vector:

$$DGD_{g_{dif}} = \sum_{r=1}^{L/2} M_{RDM}, \quad 4.21$$

$$DOD_r = \sum_{g_{dif}=0}^{G-1} M_{RDM} \cdot g_{dif}, \quad 4.22$$

$$DAD_{g_{dif}} = \sum_{r=1}^{L/2} M_{RDM} \cdot r \quad 4.23$$

From these three characteristics we can extract the following five features:

- 1- Large difference emphasis (LDE), which measures the predominance of large gray level differences.

$$x_{57} = LDE = \sum_{g_{dif}=0}^{G-1} DGD(g_{dif}) \cdot \ln(2/g_{dif}), \quad 4.24$$

- 2- Sharpness, which measures the contrast and definition in an image.

$$x_{58} = \text{Sharpness} = \sum_{g_{dif}=0}^{G-1} DGD(g_{dif}) \cdot (g_{dif})^3, \quad 4.25$$

- 3- Second Moment of DGD (SMG), which measures the variation of gray level differences.

$$x_{59} = SMG = \sum_{g_{dif}=0}^{G-1} (DGD(g_{dif}))^2, \quad 4.26$$

- 4- Second Moment of DOD (SMO), which measures the variation of average gray level differences.

$$x_{60} = SMO = \sum_{r=1}^{L/2} (DOD(r))^2, \quad 4.27$$

- 5- Long distance emphasis for large difference (LDEL), which measures the prominence of large differences a long distance from each other.

$$x_{61} = LDEL = \sum_{g_{dif}=0}^{G-1} DAD(g_{dif}) \cdot (g_{dif})^2. \quad 4.28$$

The pseudo code description of the Run Difference Matrix is given in Algorithm 4.2.

As in SGLDM algorithm, in steps 6-9 we create eight images of the origin ROI in order to extract features from each of which at one direction ($\theta=0$).

In step 16 we create the RDM matrix, and then the measure DGD, DOD and DAD which are computed in steps 21, 22, and 23. From the computed measures, we can then extract the five features (LDE; Sharpness; SMG; SMO; LDEL) as shown in step 24.

Algorithm 4.2: <i>The pseudo code of RDM technique.</i>	
Purpose:	1. Extracting 5 features from the mammogram images.
Input:	2. Enhanced segmented objects with black background (I). 3. The used direction for co-occurrence $\theta = \{0^\circ, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, 7\pi/8\}$.
Output:	4. The RDM extracted features vector of the input image I.
Procedure:	5. <u>begin</u> 6. <u>for each angle θ do</u> 7. create I_θ , which is the image rotated by angle θ . 8. $I_\theta = \text{crop } I_\theta$ to get the inscribed squares within the I_θ . 9. <u>end loop;</u> 10. $L = \text{image side};$ 11. <u>for each angle θ do</u> 12. <u>for distances $d=1$ to $L/2$ do</u> 13. <u>for height $h=1$ to L do</u> 14. <u>for pixel $p=1$ to $(L-d)$ do</u> 15. $G_{dif} = I_\theta(p, h) - I_\theta(p + d, h) + 1;$ 16. $RDM_\theta(d, G_{dif}) = RDM_\theta(d, G_{dif}) + 1;$ 17. <u>end loop;</u> 18. <u>end loop;</u> 19. <u>end loop;</u> 20. <u>end loop;</u> 21. $DGD = \text{computeDGD}; // \text{see equation 4.21}$ 22. $DOD = \text{computeDOD}; // \text{see equation 4.22}$ 23. $DAD = \text{computeDAD}; // \text{see equation 4.23}$ 24. $RDMfeatures = [LDE; \text{Sharpness}; \text{SMG}; \text{SMO}; \text{LDEL}];$ <div style="text-align: right;">$// \text{see equations}$ $// 4.24 - 4.28$</div> 25. return $RDMfeatures$ 26. <u>end</u>

4.2.3 Square Centroid Lines Gray Level Distribution Method (SCLGM)

We introduce a new extracting feature method; we call it Square Centroid Lines Gray level distribution Method (SCLGM). We use the smallest square that includes the segmented mass with zero background. As shown in Figure 4-4, the relations between the pixels of each centroid line provide discriminative information about the mass type.

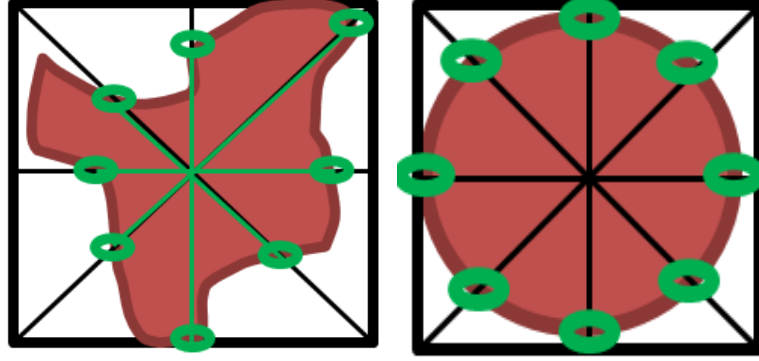


Figure 4-4: Centroid lines pixels provide discriminative information about the mass type.

In our method, we choose four centroid lines (C_i): C_1 at $\theta = 0$, C_2 at $\theta = \pi/4$, C_3 at $\theta = \pi/2$ and C_4 at $\theta = 3\pi/4$, which are the lines between the center and the vertices of the square. We represent the gray level points at each centroid line by C_{gi} :

$$C_{gi} = \left\{ \forall (x, y) : x \sin \theta - y \cos \theta - x_c \sin \theta + y_c \cos \theta = 0, \right. \\ \text{where } (x_c, y_c) \text{ is the square center,} \\ \left. (x, y) \in ROI \text{ and } \theta \in \left\{ 0, \frac{\pi}{4}, \frac{2\pi}{4}, \frac{3\pi}{4} \right\} \right\} \quad 4.29$$

After defining C_{gi} , and as we do in SGLDM, a set of statistics need to be computed before extracting textural features. In the following, we compute these statistics and extracted textural features:

1- **Mean:** which measure the average of the gray level of each line (C_{gi}):

$$\overline{C_{gi}} = \frac{1}{|C_{gi}|} \sum_{n=1}^{|C_{gi}|} C_{gi}(n) \quad 4.30$$

2- **Variance and its difference vector:** variance is used to measure the variability of the gray levels for each central line around their mean value:

$$\sigma_i = \frac{1}{|C_{gi}|} \sum_{n=1}^{|C_{gi}|} (C_{gi}(n) - \overline{C_{gi}})^2 \quad 4.31$$

As a result, we have four variance values one for each line, then we compute the differences between these variances as a vector Δ_σ :

$$\Delta_\sigma = (|\sigma_i - \sigma_j| \forall i \in \{1,2,3,4\} \text{ and } j \in \{2,3,4\}: i \neq j) \quad 4.32$$

The extracted features:

- Average variance:

$$x_{62} = \frac{1}{4} \sum_{i=1}^4 \sigma_i \quad 4.33$$

- Average difference of variance:

$$x_{63} = \frac{1}{|\Delta_\sigma|} \sum_{i=1}^{|\Delta_\sigma|} \Delta_\sigma(i) \quad 4.34$$

- Minimum variance:

$$x_{64} = \min_{i=1..4} \sigma_i \quad 4.35$$

- Minimum difference of variance:

$$x_{65} = \min_{i=1..|\Delta_\sigma|} \Delta_\sigma(i) \quad 4.36$$

- Maximum variance:

$$x_{66} = \max_{i=1..4} \sigma_i \quad 4.37$$

- Maximum difference of variance:

$$x_{67} = \max_{i=1..|\Delta_\sigma|} \Delta_\sigma(i) \quad 4.38$$

- Range of variance:

$$x_{68} = x_{66} - x_{64} \quad 4.39$$

- Range of difference of variance:

$$x_{69} = x_{65} - x_{67} \quad 4.40$$

- Median variance:

$$x_{70} = \text{median}_{i=1..4}(\sigma_i) \quad 4.41$$

- Median difference of variance:

$$x_{71} = \text{median}_{i=1..|\Delta_\sigma|}(\Delta_\sigma(i)) \quad 4.42$$

- Standard deviation of variance:

$$x_{72} = \frac{1}{4} \sum_{i=1}^4 (\sigma_i - x_{62})^2 \quad 4.43$$

- Standard deviation of difference of variance:

$$x_{73} = \frac{1}{|\Delta_\sigma|} \sum_{i=1}^{|\Delta_\sigma|} (\Delta_\sigma(i) - x_{63})^2 \quad 4.44$$

- Mean absolute deviation of variance:

$$x_{74} = \frac{1}{4} \sum_{i=1}^4 |\sigma_i - x_{62}| \quad 4.45$$

- MAD of difference of variance:

$$x_{75} = \frac{1}{|\Delta_\sigma|} \sum_{i=1}^{|\Delta_\sigma|} |\Delta_\sigma(i) - x_{63}| \quad 4.46$$

- Skewness of variance:

$$x_{76} = \frac{\frac{1}{4} \sum_{i=1}^4 (\sigma_i - x_{62})^3}{\left(\frac{1}{4} \sum_{i=1}^4 (\sigma_i - x_{62})^2 \right)^{\frac{3}{2}}}, \quad 4.47$$

- Skewness of difference of variance:

$$x_{77} = \frac{\frac{1}{|\Delta_\sigma|} \sum_{i=1}^{|\Delta_\sigma|} (\Delta_\sigma(i) - x_{63})^3}{\left(\frac{1}{|\Delta_\sigma|} \sum_{i=1}^{|\Delta_\sigma|} (\Delta_\sigma(i) - x_{63})^2 \right)^{\frac{3}{2}}}, \quad 4.48$$

3- Standard deviation (δ) and its difference vector (Δ_δ):

$$\delta_i = \sqrt[2]{\sigma_i} \quad 4.49$$

$$\Delta_\delta = \sqrt[2]{\Delta_\sigma} \quad 4.50$$

The extracted features:

- Average of Standard deviation:

$$x_{78} = \frac{1}{4} \sum_{i=1}^4 \delta_i \quad 4.51$$

- Average of difference of Standard deviation:

$$x_{79} = \frac{1}{|\Delta_\delta|} \sum_{i=1}^{|\Delta_\delta|} \Delta_\delta(i) \quad 4.52$$

- Minimum Standard deviation:

$$x_{80} = \min_{i=1..4} \delta_i \quad 4.53$$

- Minimum difference of Standard deviation:

$$x_{81} = \min_{i=1..|\Delta_\delta|} \Delta_\delta(i) \quad 4.54$$

- Maximum Standard deviation:

$$x_{82} = \max_{i=1..4} \delta_i \quad 4.55$$

- Maximum difference of Standard deviation:

$$x_{83} = \max_{i=1..|\Delta_\delta|} \Delta_\delta(i) \quad 4.56$$

- Range of Standard deviation:

$$x_{84} = x_{82} - x_{80} \quad 4.57$$

- Range of difference of Standard deviation:

$$x_{85} = x_{81} - x_{83} \quad 4.58$$

- Median Standard deviation:

$$x_{86} = \text{median}_{i=1..4}(\delta_i) \quad 4.59$$

- Median difference of Standard deviation:

$$x_{87} = \text{median}_{i=1..|\Delta_\delta|}(\Delta_\delta(i)) \quad 4.60$$

- Standard deviation of Standard deviation:

$$x_{88} = \frac{1}{4} \sum_{i=1}^4 (\delta_i - x_{78})^2 \quad 4.61$$

- Standard deviation of difference of Standard deviation:

$$x_{89} = \frac{1}{|\Delta_\delta|} \sum_{i=1}^{|\Delta_\delta|} (\Delta_\delta(i) - x_{79})^2 \quad 4.62$$

- MAD of Standard deviation:

$$x_{90} = \frac{1}{4} \sum_{i=1}^4 |\delta_i - x_{78}| \quad 4.63$$

- MAD of difference of Standard deviation:

$$x_{91} = \frac{1}{|\Delta_\delta|} \sum_{i=1}^{|\Delta_\delta|} |\Delta_\delta(i) - x_{79}| \quad 4.64$$

- Skewness of Standard deviation:

$$x_{92} = \frac{\frac{1}{4} \sum_{i=1}^4 (\delta_i - x_{78})^3}{\left(\frac{1}{4} \sum_{i=1}^4 (\delta_i - x_{78})^2 \right)^{\frac{3}{2}}}, \quad 4.65$$

- Skewness of difference of Standard deviation:

$$x_{93} = \frac{\frac{1}{|\Delta_\delta|} \sum_{i=1}^{|\Delta_\delta|} (\Delta_\delta(i) - x_{79})^3}{\left(\frac{1}{|\Delta_\delta|} \sum_{i=1}^{|\Delta_\delta|} (\Delta_\delta(i) - x_{79})^2 \right)^{\frac{3}{2}}}, \quad 4.66$$

4- **Mean absolute deviation (MAD) and its differences** : MAD is a robust measure of the variability of gray level distribution around the mean, for each centroid line i we compute the MAD as follows:

$$MAD_i = \frac{1}{|C_{gi}|} \sum_{n=1}^{|C_{gi}|} |C_{gi}(n) - \overline{C_{gi}}| \quad 4.67$$

The differences of MAD could be computed as follow:

$$\Delta_{MAD} = (|MAD_i - MAD_j| \forall i \in \{1,2,3,4\} \text{ and } j \in \{2,3,4\}: i \neq j) \quad 4.68$$

The extracted features:

- Average of MAD:

$$x_{94} = \frac{1}{4} \sum_{i=1}^4 MAD_i \quad 4.69$$

- Average of difference of MAD:

$$x_{95} = \frac{1}{|\Delta_{MAD}|} \sum_{i=1}^{|\Delta_{MAD}|} \Delta_{MAD}(i) \quad 4.70$$

- Minimum of MAD:

$$x_{96} = \min_{i=1..4} MAD_i \quad 4.71$$

- Minimum of difference of MAD:

$$x_{97} = \min_{i=1..|\Delta_{MAD}|} \Delta_{MAD}(i) \quad 4.72$$

- Maximum of MAD:

$$x_{98} = \max_{i=1..4} MAD_i \quad 4.73$$

- Maximum of difference of MAD:

$$x_{99} = \max_{i=1..|\Delta_{MAD}|} \Delta_{MAD}(i) \quad 4.74$$

- Range of MAD:

$$x_{100} = x_{98} - x_{96} \quad 4.75$$

- Range of difference of MAD:

$$x_{101} = x_{97} - x_{99} \quad 4.76$$

- Median of MAD:

$$x_{102} = median_{i=1..4}(MAD_i) \quad 4.77$$

- Median difference of MAD:

$$x_{103} = median_{i=1..|\Delta_{MAD}|}(\Delta_{MAD}(i)) \quad 4.78$$

- Standard deviation of MAD:

$$x_{104} = \frac{1}{4} \sum_{i=1}^4 (MAD_i - x_{94})^2 \quad 4.79$$

- Standard deviation of difference of MAD:

$$x_{105} = \frac{1}{|\Delta_{MAD}|} \sum_{i=1}^{|\Delta_{MAD}|} (\Delta_{MAD}(i) - x_{95})^2 \quad 4.80$$

- Mean absolute deviation of MAD:

$$x_{106} = \frac{1}{4} \sum_{i=1}^4 |MAD_i - x_{94}| \quad 4.81$$

- Mean absolute deviation of difference of MAD:

$$x_{107} = \frac{1}{|\Delta_{MAD}|} \sum_{i=1}^{|\Delta_{MAD}|} |\Delta_{MAD}(i) - x_{95}| \quad 4.82$$

- Skewness of MAD:

$$x_{108} = \frac{\frac{1}{4} \sum_{i=1}^4 (MAD_i - x_{94})^3}{\left(\frac{1}{4} \sum_{i=1}^4 (MAD_i - x_{94})^2 \right)^{\frac{3}{2}}} \quad 4.83$$

- Skewness of difference of MAD:

$$x_{109} = \frac{\frac{1}{|\Delta_{MAD}|} \sum_{i=1}^{|\Delta_{MAD}|} (\Delta_{MAD}(i) - x_{95})^3}{\left(\frac{1}{|\Delta_{MAD}|} \sum_{i=1}^{|\Delta_{MAD}|} (\Delta_{MAD}(i) - x_{95})^2 \right)^{\frac{3}{2}}} \quad 4.84$$

- 5- **Skewness (SK)**: skewness is a measure of the asymmetry of the gray level distribution, it is computed by the following formula:

$$SK_i = \frac{\frac{1}{|C_{gi}|} \sum_{n=1}^{|C_{gi}|} (C_{gi}(n) - \overline{C_{gi}})^3}{\left(\frac{1}{|C_{gi}|} \sum_{n=1}^{|C_{gi}|} (C_{gi}(n) - \overline{C_{gi}})^2 \right)^{\frac{3}{2}}} \quad 4.85$$

Skewness differences could be computed as follow:

$$\Delta_{SK} = (|SK_i - SK_j| \forall i \in \{1,2,3,4\} \text{ and } j \in \{2,3,4\}: i \neq j) \quad 4.86$$

The extracted features:

- Average of SK:

$$x_{110} = \frac{1}{4} \sum_{i=1}^4 SK_i \quad 4.87$$

- Average of difference of SK:

$$x_{111} = \frac{1}{|\Delta_{SK}|} \sum_{i=1}^{|\Delta_{SK}|} \Delta_{SK}(i) \quad 4.88$$

- Minimum of SK:

$$x_{112} = \min_{i=1..4} SK_i \quad 4.89$$

- Minimum of difference of SK:

$$x_{113} = \min_{i=1..|\Delta_{SK}|} \Delta_{SK}(i) \quad 4.90$$

- Maximum of SK:

$$x_{114} = \max_{i=1..4} SK_i \quad 4.91$$

- Maximum of difference of SK:

$$x_{115} = \max_{i=1..|\Delta_{SK}|} \Delta_{SK}(i) \quad 4.92$$

- Range of SK:

$$x_{116} = x_{114} - x_{112} \quad 4.93$$

- Range of difference of SK:

$$x_{117} = x_{113} - x_{115} \quad 4.94$$

- Median of SK:

$$x_{118} = median_{i=1..4}(SK_i) \quad 4.95$$

- Median of difference of SK:

$$x_{119} = median_{i=1..|\Delta_{SK}|}(\Delta_{SK}(i)) \quad 4.96$$

- Standard deviation of SK:

$$x_{120} = \frac{1}{4} \sum_{i=1}^4 (SK_i - x_{110})^2 \quad 4.97$$

- Standard deviation of difference of SK:

$$x_{121} = \frac{1}{|\Delta_{SK}|} \sum_{i=1}^{|\Delta_{SK}|} (\Delta_{SK}(i) - x_{111})^2 \quad 4.98$$

- Mean absolute deviation of SK:

$$x_{122} = \frac{1}{4} \sum_{i=1}^4 |SK_i - x_{110}| \quad 4.99$$

- Mean absolute deviation of difference of SK:

$$x_{123} = \frac{1}{|\Delta_{SK}|} \sum_{i=1}^{|\Delta_{SK}|} |\Delta_{SK}(i) - x_{111}| \quad 4.100$$

- Skewness of SK:

$$x_{124} = \frac{\frac{1}{4} \sum_{i=1}^4 (SK_i - x_{110})^3}{\left(\frac{1}{4} \sum_{i=1}^4 (SK_i - x_{110})^2 \right)^{\frac{3}{2}}}, \quad 4.101$$

- Skewness of difference of SK:

$$x_{125} = \frac{\frac{1}{|\Delta_{SK}|} \sum_{i=1}^{|\Delta_{SK}|} (\Delta_{SK}(i) - x_{111})^3}{\left(\frac{1}{|\Delta_{SK}|} \sum_{i=1}^{|\Delta_{SK}|} (\Delta_{SK}(i) - x_{111})^2 \right)^{\frac{3}{2}}}, \quad 4.102$$

- 6- **Kurtosis (K):** Kurtosis is a measure of whether the gray level distribution is peaked or flat relative to a normal distribution.

$$K_i = \frac{\frac{1}{|C_{gi}|} \sum_{n=1}^{|C_{gi}|} (C_{gi}(n) - \overline{C_{gi}})^4}{\left(\frac{1}{|C_{gi}|} \sum_{n=1}^{|C_{gi}|} (C_{gi}(n) - \overline{C_{gi}})^2 \right)^2} - 3 \quad 4.103$$

Kurtosis differences vector could be computed as follows:

$$\Delta_K = (|K_i - K_j| \forall i \in \{1,2,3,4\} \text{ and } j \in \{2,3,4\}; i \neq j) \quad 4.104$$

The extracted features:

- Average of Kurtosis:

$$x_{126} = \frac{1}{4} \sum_{i=1}^4 K_i \quad 4.105$$

- Average of difference of Kurtosis:

$$x_{127} = \frac{1}{|\Delta_K|} \sum_{i=1}^{|\Delta_K|} \Delta_K(i) \quad 4.106$$

- Minimum of Kurtosis:

$$x_{128} = \min_{i=1..4} K_i \quad 4.107$$

- Minimum of difference of Kurtosis:

$$x_{129} = \min_{i=1..|\Delta_K|} \Delta_K(i) \quad 4.108$$

- Maximum of Kurtosis:

$$x_{130} = \max_{i=1..4} K_i \quad 4.109$$

- Maximum of difference of Kurtosis:

$$x_{131} = \max_{i=1..|\Delta_K|} \Delta_K(i) \quad 4.110$$

- Range of Kurtosis:

$$x_{132} = x_{130} - x_{128} \quad 4.111$$

- Range of difference of Kurtosis:

$$x_{133} = x_{129} - x_{131} \quad 4.112$$

- Median of Kurtosis:

$$x_{134} = \text{median}_{i=1..4}(K_i) \quad 4.113$$

- Median of difference of Kurtosis:

$$x_{135} = \text{median}_{i=1..|\Delta_K|}(\Delta_K(i)) \quad 4.114$$

- Standard deviation of Kurtosis:

$$x_{136} = \frac{1}{4} \sum_{i=1}^4 (K_i - x_{126})^2 \quad 4.115$$

- Standard deviation of difference of Kurtosis:

$$x_{137} = \frac{1}{|\Delta_K|} \sum_{i=1}^{|\Delta_K|} (\Delta_K(i) - x_{127})^2 \quad 4.116$$

- Mean absolute deviation of Kurtosis:

$$x_{138} = \frac{1}{4} \sum_{i=1}^4 |K_i - x_{126}| \quad 4.117$$

- Mean absolute deviation of difference of Kurtosis

$$x_{139} = \frac{1}{|\Delta_K|} \sum_{i=1}^{|\Delta_K|} |\Delta_K(i) - x_{127}| \quad 4.118$$

- Skewness of Kurtosis:

$$x_{140} = \frac{\frac{1}{4} \sum_{i=1}^4 (K_i - x_{126})^3}{\left(\frac{1}{4} \sum_{i=1}^4 (K_i - x_{126})^2 \right)^{\frac{3}{2}}} \quad 4.119$$

- Skewness of difference of Kurtosis:

$$x_{141} = \frac{\frac{1}{|\Delta_K|} \sum_{i=1}^{|\Delta_K|} (\Delta_K(i) - x_{127})^3}{\left(\frac{1}{|\Delta_K|} \sum_{i=1}^{|\Delta_K|} (\Delta_K(i) - x_{127})^2 \right)^{\frac{3}{2}}} \quad 4.120$$

The pseudo code description of the proposed SCLGM technique is given in Algorithm 4.3.

Algorithm 4.3: <i>The pseudo code of the proposed SCLGM technique.</i>	
Purpose:	1. Extracting 75 features from the mammogram images.
Input:	2. Enhanced segmented objects with black background (I). 3. The used direction for co-occurrence $\theta = \{0^\circ, \pi/4, \pi/2, 3\pi/4\}$.
Output:	4. The SCLGM extracted features vector of the input image I.
Procedure:	5. <u>begin</u> 6. <u>for each angle θ do</u> 7. $l = 0$; 8. $c = \text{floor}((L + 1)/2)$; //image side center 9. <u>for points $x=-c$ to c do</u> 10. $y = \tan(\theta) * x$; 11. if $I_\theta(x, y) \neq 0$ then 12. $l = l + 1$; 13. $C_\theta[l] = I_\theta(x, y)$; 14. end if; 15. <u>end loop</u> ; 16. $M_{\theta 1} = \text{mean}(C_\theta)$; //Compute the following measures 17. $M_{\theta 2} = \text{variance}(C_\theta)$; 18. $M_{\theta 3} = \text{varianceDifference}(C_\theta)$; 19. $M_{\theta 4} = \text{standardVariation}(C_\theta)$; 20. $M_{\theta 5} = \text{standardVariationDifference}(C_\theta)$; 21. $M_{\theta 6} = \text{MAD}(C_\theta)$; 22. $M_{\theta 7} = \text{MADDifference}(C_\theta)$; 23. $M_{\theta 8} = \text{skewness}(C_\theta)$; 24. $M_{\theta 9} = \text{skewnessDifference}(C_\theta)$; 25. $M_{\theta 10} = \text{kurtosis}(C_\theta)$; 26. $M_{\theta 11} = \text{kurtosisDifference}(C_\theta)$; 27. <u>end loop</u> ; 28. <u>for each measure vector M_i do</u> 29. $\text{SCLGM}_{\text{Features}} = \text{extract}(\text{average}(M_i), \text{minimum}(M_i),$ $\text{maximum}(M_i), \text{range}(M_i), \text{median}(M_i), \text{standarddeviation}(M_i),$ $\text{MAD}(M_i), \text{skewness}(M_i))$; 30. end loop; 31. return $\text{SCLGM}_{\text{Features}}$; 32. end;

The mentioned four centroid lines are extracted and represented by four vectors C_0 as detailed in steps 9-15 of Algorithm 4.3, each vector contains all points with nonzero gray level value. After extracting the vectors of centroid lines, we compute 11 measures for each vector as shown in steps 16-26. Then we extract the SCLGM features in 28-30 steps.

4.2.4 Human Features

In addition to the computational extracting features, we use human extracting features, which are given by DDSM [46]. The interpretation description of these human features are based on Breast Imaging Reporting and Data System (BI-RADS) [3], which are used as a guide for standardizing mammographic reporting. In the following, we describe in brief using BI-RADS the used human extracted features:

$$x_{142} = \text{patient age}, \quad 4.121$$

$$x_{143} = \text{Density}, \quad 4.122$$

We use values between 1 and 4 to describe the breast density, the difficulties increased as the value increased. These values are interpreted with respect to BI-RADS [3] definition.

Table 4-1: BI-RADS interpretation for breast density values.

Density value	BI-RADS interpretation
1	The breast is almost entirely fat.
2	There are scattered fibroglandular densities that could obscure a lesion on mammography.
3	The breast is heterogeneously dense. This may lower the sensitivity of mammography.
4	The breast tissue is extremely dense, which lowers the sensitivity of mammography.

$$x_{144} = \text{mass shape}, \quad 4.123$$

The mass shape is described by either one or more of the following five keywords:

Round: a mass that is spherical, ball-shaped, circular or globular in shape.

Oval: a mass that is elliptical or egg-shaped.

Lobulated: a mass that has contours with undulations.

Irregular: the lesion's shape cannot be characterized by any of the above.

Architectural distortion: The normal architecture is distorted with no definite mass visible.

and $x_{145} = \text{mass margins}$. 4.124

The mass margin is described by either one or more of the following five keywords:

Circumscribed: the margins are sharply demarcated with an abrupt transition between the lesion and the surrounding tissue. Without additional modifiers there is nothing to suggest infiltration.

Microlobulated: the margins undulate with short cycles producing small undulations.

Obscured: one which is hidden by superimposed or adjacent normal tissue and cannot be assessed any further.

Ill defined: The poor definition of the margins raises concern that there may be infiltration by the lesion and this is not likely due to superimposed normal breast tissue.

Spiculated: The lesion is characterized by lines radiating from the margins of a mass.

4.3 Feature Selection

The main goal of feature selection is to reduce the dimensionality by eliminating irrelevant features and selecting the best discriminative features. Many search methods are proposed for feature selection. As mentioned before, these methods could be categorized into sequential or randomized feature selection. Sequential methods are simple and fast but they could not backtrack, which means that they are candidate to fall into local minima. The problem of local minima is solved in randomized feature selection methods, but with randomized methods it is difficult to choose proper parameters.

To avoid problems of local minima and choosing proper parameters, we use two methods for feature selection, the sequential and the genetic algorithm (GA) feature selection. We use all extracted features as input for both sequential feature selection and GA feature selection individually.

In forward sequential method, we start with empty list of selected feature, and successively we add one useful feature to the list until no useful feature remains in the extracted input list. The selection of useful feature is based on a criteria function, which could be a correlation, a distance, a misclassification error or an entropy measure. Our criteria function based on misclassification error is defined as follows:

$$E = \frac{1}{N} \sum_{d=1}^N I(o_d \neq t_d) \quad 4.125$$

Where I is the image to be classified (training example), N is the number of training examples, o_d and t_d are the output value and the target value for training example d .

We choose SVM classification method in criteria function, that SVM training always finds a global minimum.

In GA algorithm, we use a fitness function based on the principle of Max-Relevance and Min-Redundancy (mRMR) [47] to have a discriminant power feature with minimum redundancy. The idea of mRMR is to select set S with m features $\{x_i\}$ that satisfies the maximization problem:

$$\max_i \Phi_i(D, R), \quad \Phi(D, R) = D - R \quad 4.126$$

Where D and R represent the max- relevance and min-redundancy, respectively, and are computed by the following formula:

$$D = \frac{1}{|S|} \sum_{x_i} I(x_i; y) \quad 4.127$$

$$R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \quad 4.128$$

where $I(x_i; y)$ and $I(x_i, x_j)$ represent the mutual information, which is the quantity that measures the mutual dependence of the two random variables and is computed using the following formula-see Figure 4-5-:

$$I(x; y) = H(x) + H(y) - H(x, y) \quad 4.129$$

where $H(.)$ is the entropy of the respective multivariate or univariate variables.

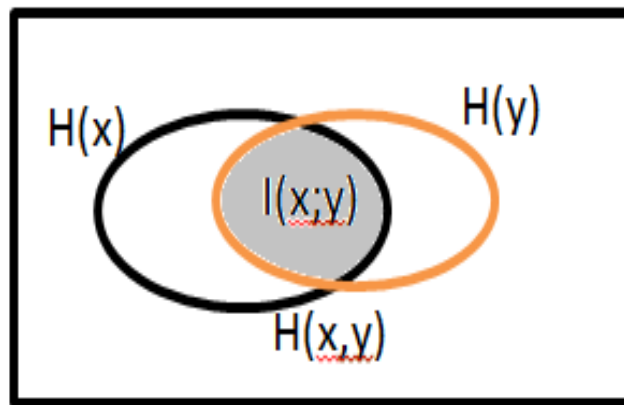


Figure 4-5: The mutual information $I(x; y)$ is the intersection of the entropies of each x and y

As shown in figure 4.6, we apply feature selection methods 10 times, each time the results are slightly changed from previous. Then, the results are combined and ordered with respect to both the number of occurrences and the accuracy values that are obtained by criteria or fitness function. Algorithm 4.4 describes the steps of feature selection.

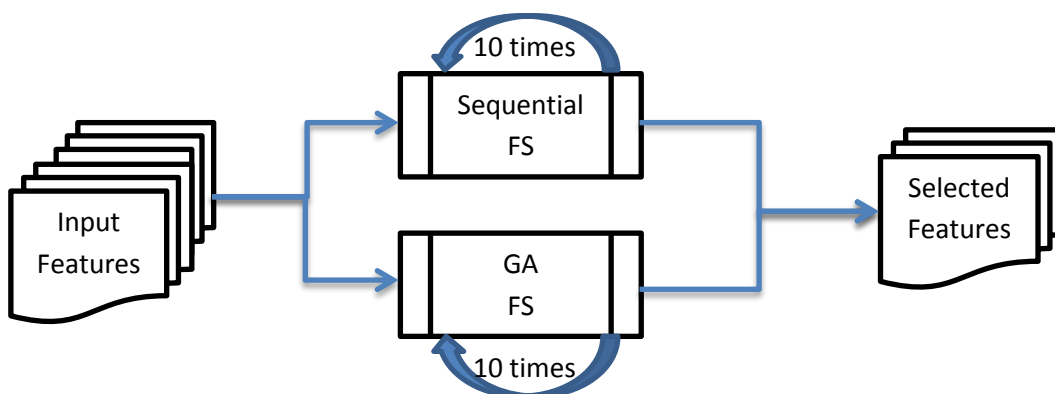


Figure 4-6: Feature selection technique.

As shown in Algorithm 4.4, the selected features are represented by four sets:

- 1- Sequential Forward Features (SFF): which are the features selected by Sequential Forward Selection technique.
- 2- Genetic Algorithm Features (GAF): which are the features selected by Genetic Algorithm technique.
- 3- Union of Selected Features (USF): USF contains the union of features of SFF and GAF.
- 4- Intersection of Selected Features (ISF): represents the shared features between SFF and GAF.

Algorithm 4.4: <i>The pseudo code of the proposed feature selection technique.</i>	
Purpose:	1. Selecting the best and discriminative features from the extracted set.
Input:	2. 145 extracted features.
Output:	3. The selected features are represented by the following sets: <i>SFF</i> , <i>GAF</i> , <i>USF</i> , and <i>ISF</i> .
Procedure:	4. <u>begin</u> 5. <u>for</u> $i = 1$ to 10 <u>do</u> 6. $SFF =$ $UNION(SFF, \text{Features selected by Sequential Forward});$ 7. // see algorithm 3.1 8. $GAF = UNION(GAF, \text{Features selected by GA});$ 9. // see algorithm 3.2 10. <u>end loop;</u> 11. $USF = UNION(SFF, GAF);$ 12. $ISF = INTERSECT(SFF, GAF);$ 13. return $SFF, GAF, USF, ISF;$ 14. <u>end;</u>

Chapter 5

Experimentation and Results

5.1 DDSM Database

The Digital Database for Screening Mammography (DDSM) is a resource for use by the mammographic image analysis research community. It is maintained at the University of South Florida for purposes of keeping it accessible on the web. DDSM is organized into 43 volumes, each volume is a collection of cases, and it is categorized into one of four types:

- Normal: A normal volume contains normal cases that are formed from a previous normal screening exam for a patient with a normal exam at least four years later.
- Cancers: this volume contains cancer cases formed from screening exams in which at least one pathology proven cancer was found.
- Benign: Benign cases are formed from screening exams in which something suspicious was found, but was determined not to be malignant.
- Benign without callback (BWC): that has benign cases in which no additional films or biopsy was done to make the benign finding.

In total, DDSM provides 2620 cases. Each case includes:

- 1) Four images compressed with lossless JPEG encoding -two images for each breast which covers CranioCaudal view and Mediolateral Oblique view-.
- 2) The patient's information; age at time of study, American College of Radiology (ACR) breast density rating, subtlety rating for abnormalities and ACR keyword description of abnormalities.
- 3) Image's information; scanner, spatial resolution, etc.

We download 610 mammographic images from DDSM for three datasets which are described in detail in Table 5-1.

Figure 5-1 shows three cases representing three types of mammogram: normal, benign and cancer, downloaded from DDSM. The mammogram cases are saved in Lossless JPEG format which are not readable by MATLAB, this is the first problem we faced.

More details about this problem and its solution are presented in the following subsection.

Table 5-1: The experiments are done with three datasets.

Dataset	# of Cancers	# of Benign	# of Normal	Totals
Training dataset	126	105	179	410
Testing dataset-1	35	42	23	100
Testing dataset-2	60	14	26	100

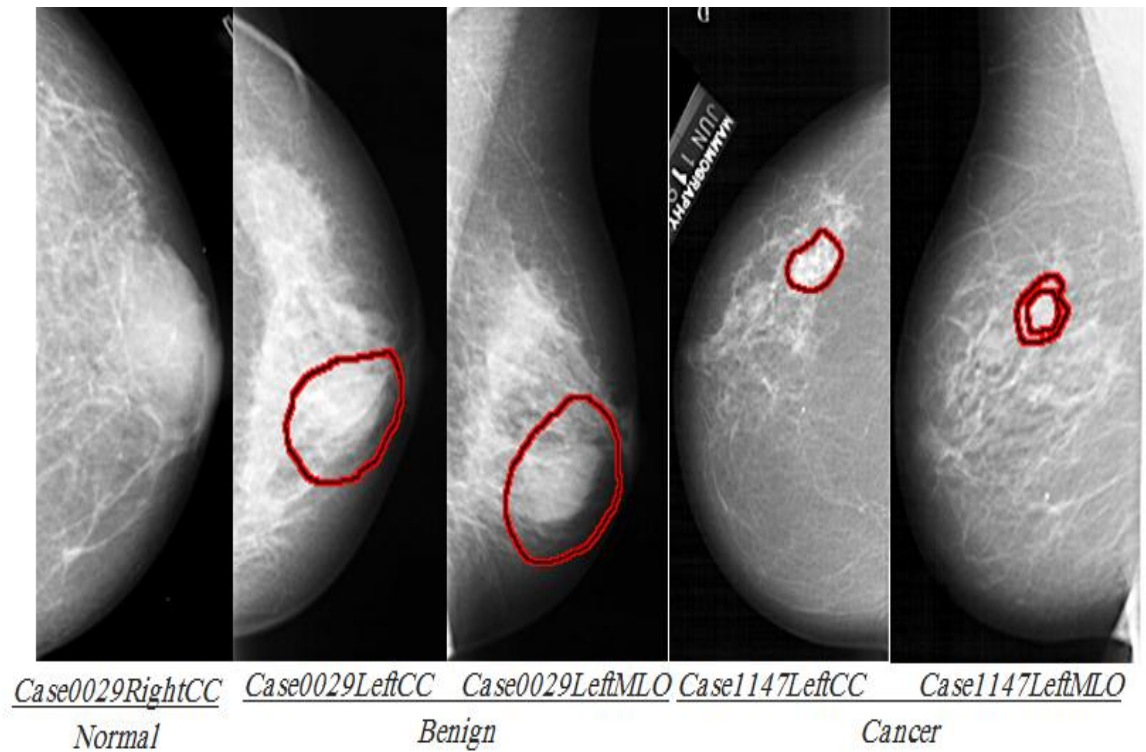


Figure 5-1: An example for each normal, benign and cancer.

5.2 Converting LJPEG to TIF

JPEG is an image compression format typically used on the web for its reduced size. In JPEG some information is discarded during compression in order to obtain smaller image file sizes. Although this lost information may be not noticeable by the human eye, it leads to reduce the quality of the image, and it can be harmful in the critical fields as in medical, military and space imaging, high-end film, professional studio-quality photography. To avoid loss of image information, DDSM use the Lossless

JPEG (LJPEG) image format – lossless means that there is no quality loss due to compression - to represent the mammography images.

LJPEG format is not readable by MATLAB, so we have to convert LJPEG mammographic images obtained from the DDSM into another lossless compression format that is readable by MATLAB. One of the most popular lossless compression image formats is TIF –Tag Image File format.

5.3 Implementation Environment

As mentioned before, we downloaded our datasets from DDSM, DDSM using LJPEG image format, which is not readable by Matlab, so we have to convert LJPEG to TIF format, which is lossless and readable by Matlab. To obtain this, we use heathusf software version 1.1.0 with gcc version 4.4.3. We run the conversion software using Oracle VM VirtualBox Version 4.0.8 r71778, and Ubuntu operating system version 11.04 - the Natty Narwhal.

The mammogram computer-aided diagnosis is implemented on windows 7 operating system using MATLAB R2011, with Matlab image processing tools and statistical tools. All experiments are implemented on a dell laptop of Intel Core 2 Due Processing power of 2.4 GHz CPU with 4GB RAM.

5.4 Manual Segmentation

We apply segmentation in two stages, the first stage is done manually and before enhancement step, the second stage is done after enhancement using gray level growing algorithm. There are two objectives for using manual segmentation:

- 1) Decreasing the image size since TIF files of mammography images are generally pretty large, so this step leads to decrease the time required for both enhancement and auto segmentation steps.
- 2) Decreasing the size of ROI for mass detection, this makes the task of auto-segmentation step easy.

As shown in Figure 5-1, the suspicious regions –benign or malignant- are marked by experienced radiologists. We use Adop-Photoshop to locate and cut the smallest

square that contains the marked suspicious region. Figure 5-2 shows some examples of manual segmentation results.

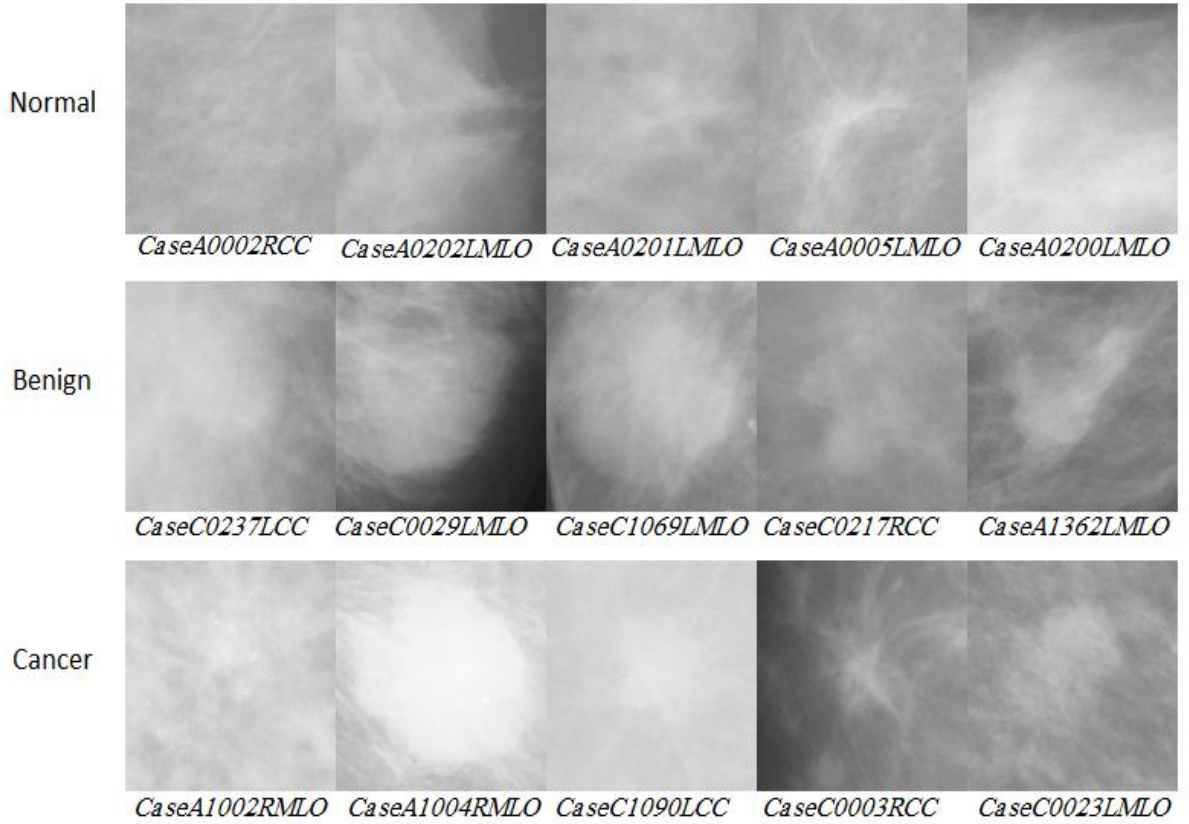


Figure 5-2: Some examples of manual segmentation.

5.5 Enhancement

It is important to do enhancement on mammogram images before applying auto segmentation for mass objects. The parenchyma density hides the mass objects and it will be so difficult to apply auto segmentation algorithms, so we use histogram equalization to enhance the mammogram images.

The main idea of histogram equalization is to use all available gray levels as shown in Figure 5-3 the histogram of original mammogram - case “C109LeftCC”- not using all available gray levels. To perform histogram equalization, we first create Probability Density Function (PDF) and the Cumulative Density Function (CDF) of mammogram image. Then we use PDF as a match to the output enhancement image. The output is shown in Figure 5-3, it is clear that after enhancement all gray levels are used.

Figure 5-4 shows some examples of mammogram images enhancements, as shown, after enhancements the mass margin are obvious and it became easy to be auto-segmented.

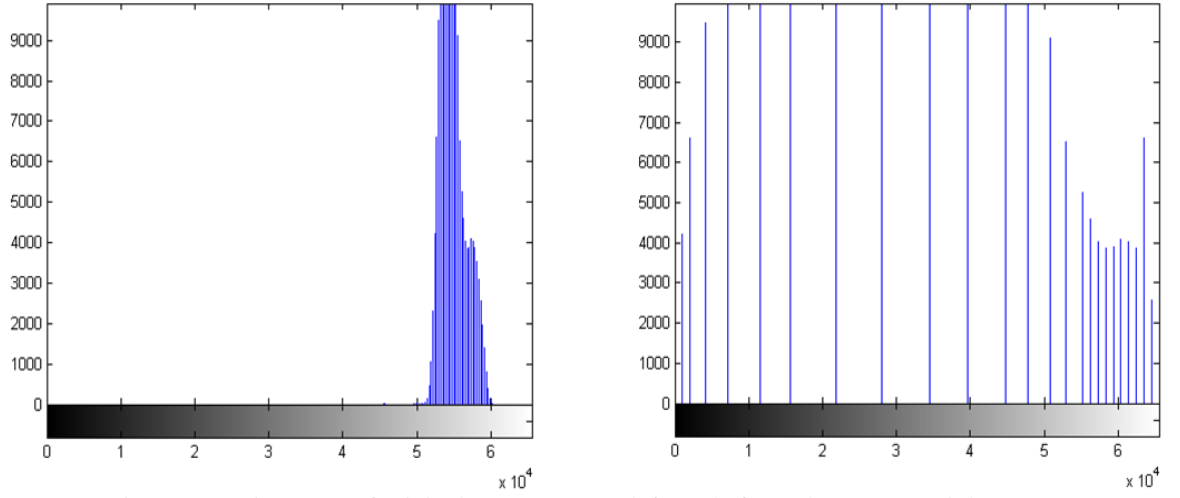


Figure 5-3: Histogram of original mammogram (left) and after enhancement (right).

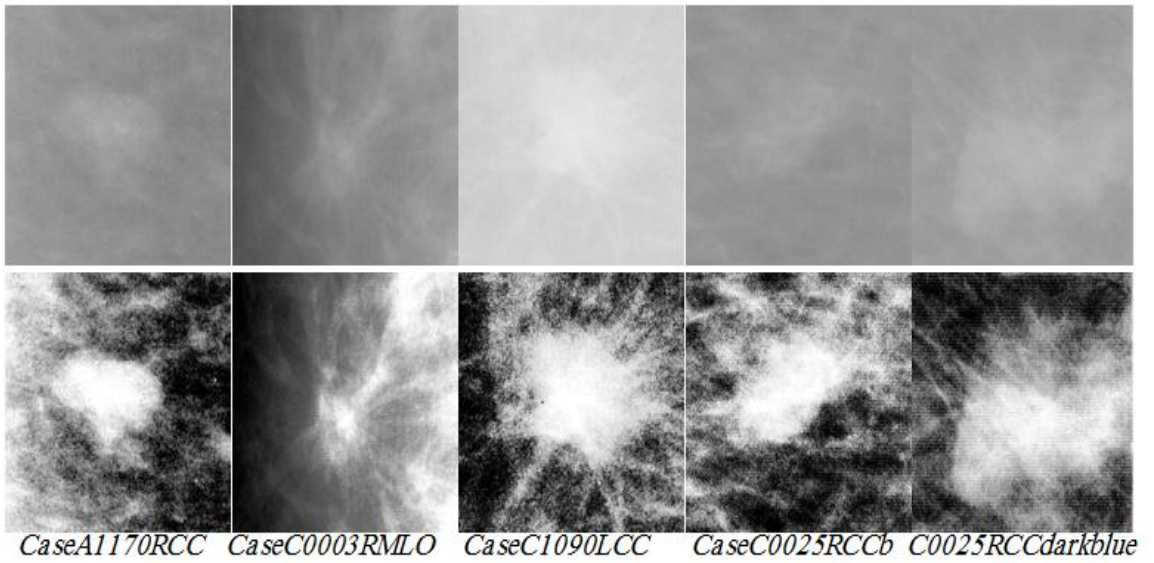


Figure 5-4: Mammogram images enhancements.

5.6 Segmentation

After mammogram enhancement, it became easy to start auto segmentation for the suspicious regions, which has an observed contour. We use region grow algorithm based on active contour segmentation. The basic idea in active contour models or snakes is to deform an initial curve so that it separates foreground from background based on the means of the two regions in order to detect objects in the processed

image. For instance, starting with a curve around the object to be detected, the curve moves toward its interior normal and has to stop on the boundary of the object [48]

This technique is very robust and gives very nice results as there is a difference between the foreground and background means. The main problem we are faced with is in selecting the seed or the curve location, at which the algorithm starts deforming the curve until locating the mass as to be a segment. This problem is solved for a while by using the curve around the center of the suspicious region, but this solution is not good if we have multiple separated masses in one region. So, we use multiple seed, distributed on a different location in the mammogram image.

Another problem we are faced with is the number of iterations required for deforming the curve in order to locate the mass. We use large numbers for iterations with possibility to stop when no changes occurred on the curve shape.

As a result, we have for each mammogram a binary mask: zero for black and one for white. By multiplying the mask image with the origin image we obtain the segmented mass as shown in Figure 5-5.

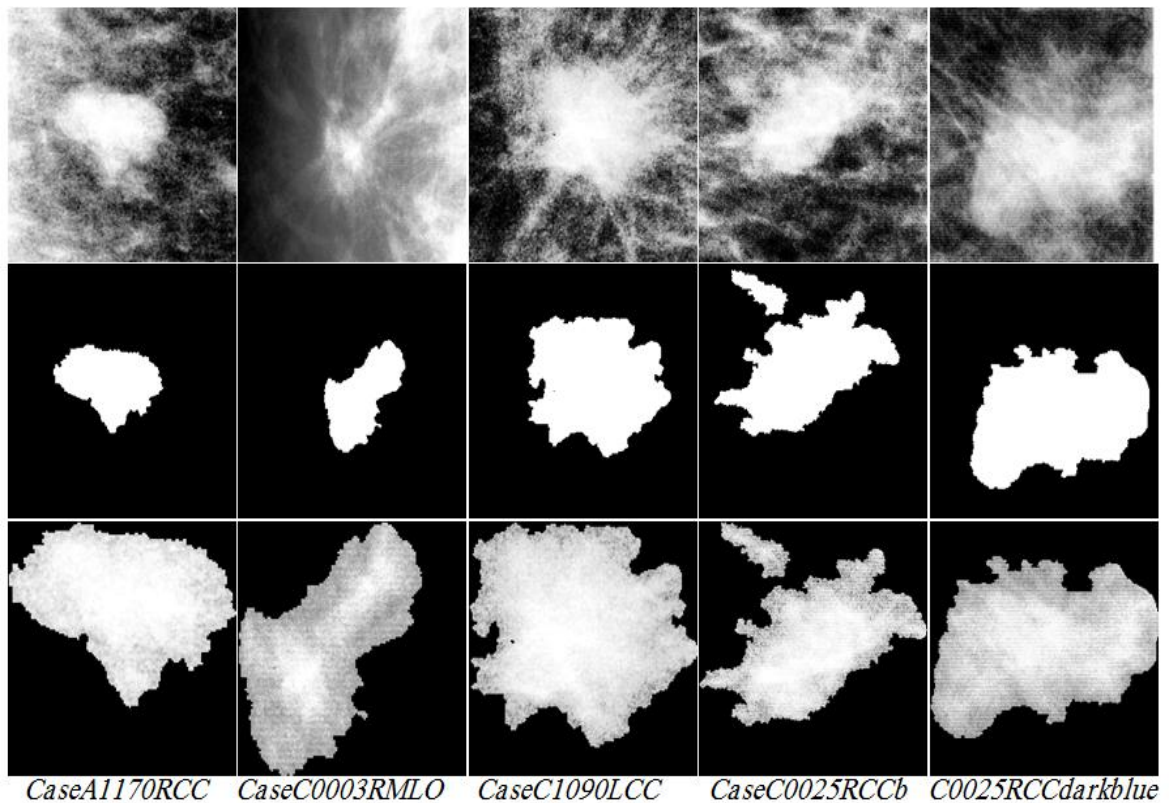


Figure 5-5: By multiplying the enhanced image (first raw) with the mask (second raw) we obtain the segmented mass shown in the third row.

5.7 Feature Extraction

As mentioned before, we have 145 features, some of these features are obtained from DDSM database –human features- and the others are extracted mathematically. The implementation of these mathematical features extraction is done by MATLAB tool using three functions:

- SGLDM_B, which is used in computing Co-occurrence matrix features.
- RDM_B, which is used in extracting the run difference matrix features.
- SCLGM_B, which is used in extracting square centroid lines gray level distribution method features.

The results of these functions are exported into excel for ordering and arranging. As a result we have a matrix of 410 rows -representing the observations of training dataset- and 145 columns -representing the extracted features.

The humanity features of some observations shown are in Table 5-2 as an example; the values of these features are described in Table 5-3, for more detail see section 4.1.4.

Table 5-2: Humanity features for some observations.

		Humanity features			
Case	Class	Age	Density	Shape	Margins
C0235RightCC	2	42	2	2	2
C0237LeftCC	2	62	3	2	1
C0029LeftCC	2	66	3	2	4
C0033RightMLO	2	60	3	2	2
C0217RightMLO	2	56	2	1	3
C1069LeftCC	1	70	2	1	5
A1357RightMLO	2	76	4	2	4
A1361LeftCC	2	54	4	4	3
A1362LeftCC	2	58	4	4	4
A1363RightMLO	2	43	4	2	1

Table 5-3: Class, margin and shape values BI-RADS description

Class		Margin		Shape	
Value	Description	Value	Description	Value	Description
1	Cancer	1	CIRCUMSCRIBED	1	Round
2	Benign	2	MICROLOBULATED	2	Oval
3	Normal	3	OBSCURED	3	Lobulated
-	-	4	ILL_DEFINED	4	Irregular
-	-	5	SPICULATED	5	Architectural_distortion

Table 5-4 shows six mathematically features extracted from contrast measure as an example of computational features.

Table 5-4: Contrast extraction features as an example of computational features.

Case	Class	Contrast					
		Mean	Min	Max	Variance	MAD	Skewness
C0235RightCC	2	3.598333	0.16129	5.89458	2.445461	1.31553	-0.4607
C0237LeftCC	2	2.250772	0.208583	4.336476	1.501606	1.062147	0.067471
C0029LeftCC	2	2.691163	0.116415	4.966352	1.762171	1.128663	-0.13232
C0033RightMLO	2	6.579114	0.181679	12.9252	12.27602	2.985507	0.02769
C0217RightMLO	2	10.84879	0.274611	16.54152	26.10129	4.437016	-0.51599
C1069LeftCC	1	2.496023	0.160451	4.067136	1.236998	0.961038	-0.40626
A1357RightMLO	2	8.423769	0.21376	15.90158	20.20636	3.854733	-0.09435
A1361LeftCC	2	7.573637	0.097616	15.04157	19.87117	3.889167	0.057665

5.8 Feature Selection

In forward sequential method we use the criteria function that is based on misclassification error. To compute misclassification error, we choose SVM classification algorithm -for its simplicity and high possibility of finding a global minimum- as a classifier in the criteria function.

In MATLAB, SVM functions support only two classes based on the basic algorithm of SVM which is designed as a binary classifier. So, we have here a problem that in our datasets we have three classes: cancer, benign and normal, so how could we extend a binary classifier SVM to multiclass? We use the idea demonstrated in Figure 5-6. We divide the three classes into two classes: “Cancer” and “Not Cancer”. If the decision of SVM choose the class “Cancer” then stop and store the result. Otherwise, we use the result as input to SVM classifier to classify it into “Benign” or “Normal”.

In our experiments, we use 10-fold cross-validation, by which the data set is divided into 10 subsets, one of them is used as a test and the remaining subsets are used for training. Repeatedly, we use a criterion function based on misclassification error for each training subset to train or fit a model, which is used next to predict the target of test subset.

The criterion function returns the sum of loss predicted targets of the test subset. Then we implement cross-validation calculation for a given candidate feature set by summing the values returned by criterion function and divide that sum by the total number of test observations. We use that mean value to evaluate each candidate feature subset.

We run forward sequential feature selection for ten times, the results of these iterations are shown in Table 5-5.

As mentioned before we use GA as a randomized feature selection method, which prevents falling in local minima. We use the following parameters:

- The population size (P) which is the number of chromosomes in each generation; we use a random value for population size computed by the following Matlab formula:

$$P = \text{round}((L-1) * \text{rand}(DF, 200 * DF)) + 1; \quad 5.1$$

Where L is the number of input features, DF is the desired number of selected features, we use its value in range of 10 to 15.

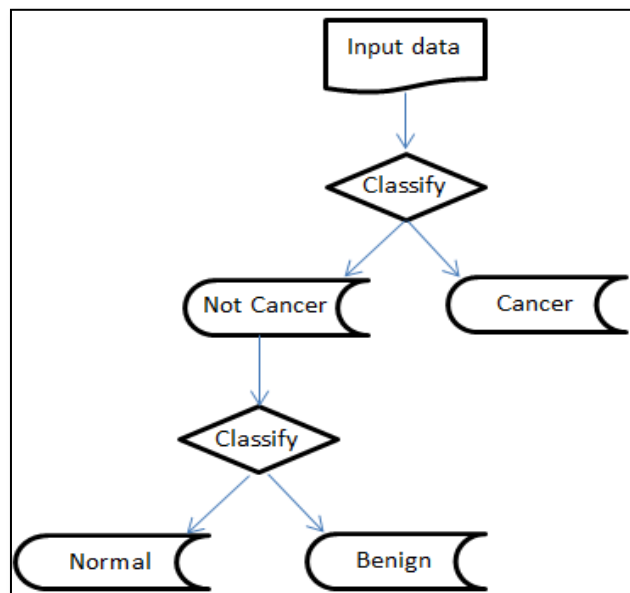


Figure 5-6: Use SVM for three classes

- Max_Generations = 200, Max_Generations stands for the maximum number of generations and is used as a termination criterion. The GA will be stopped before the generated number of chromosome populations exceeds the

Max_Generations, and the top scoring chromosome will be returned as the search's answer;

- The generations with no change in highest-scoring (elite), this parameter is used as a second termination criterion. It represents the number of generations that may pass with no changes in the elite chromosome before that elite chromosome will be returned to as the search's answer. In our method we set elite equal to one.

Table 5-5: The results of sequential forward feature selection.

Round(R) Feature	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
x_{16}	1	-	-	-	-	-	-	-	-	-
x_{22}	-	-	-	-	-	-	-	1	-	-
x_{33}	-	-	-	-	-	-	-	-	1	-
x_{39}	1	-	-	-	-	-	-	-	-	-
x_{43}	-	-	-	-	-	-	-	-	1	-
x_{44}	1	-	-	-	-	-	-	-	-	-
x_{59}	1	-	-	-	-	-	-	1	1	1
x_{60}	1	-	-	-	-	-	-	-	-	-
x_{64}	-	-	1	-	-	-	1	-	-	-
x_{69}	-	-	-	-	-	-	-	-	1	-
x_{76}	-	-	-	1	1	1	-	1	-	1
x_{80}	-	1	1	-	-	-	-	-	-	-
x_{82}	-	-	-	-	1	-	-	-	-	-
x_{84}	-	-	-	-	-	-	1	-	-	-
x_{96}	-	1	-	-	-	-	-	-	-	-
x_{98}	-	-	-	-	-	1	-	-	-	-
x_{103}	-	-	-	1	-	-	-	-	-	-
x_{114}	-	-	-	-	-	-	-	-	1	-
x_{120}	-	-	-	-	1	-	-	-	1	-
x_{126}	-	-	-	-	-	-	-	-	1	-
x_{128}	-	-	-	-	-	-	-	1	-	-
x_{135}	-	1	1	-	1	1	1	1	-	1
x_{138}	-	-	-	-	-	1	1	-	-	-
x_{139}	-	1	1	-	-	-	-	-	-	1
x_{142}	-	1	1	-	-	1	1	1	-	-
x_{143}	-	1	1	1	1	1	1	1	1	1
x_{145}	1	1	1	1	1	1	1	1	1	1

We apply GA methods 10 times for feature selection; Table 5-6 shows the selected feature for each iteration.

Table 5-6: The results of GA feature selection.

Round(R) Feature	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
x_{31}	1	1	1	1	1	1	1	1	1	1
x_{33}	1	1	1	1	1	1	1	1	1	1
x_{43}	1	1	1	1	1	1	1	1	1	1
x_{59}	1	1	1	1	1	1	1	1	1	1
x_{61}	1	1	1	1	1	1	1	1	1	1
x_{64}	1	1	1	1	1	1	1	1	1	1
x_{76}	1	-	1	1	-	-	-	-	1	1
x_{96}	1	1	1	1	1	1	1	1	1	1
x_{103}	-	-	-	-	1	1	-	-	-	-
x_{120}	-	-	-	-	1	1	-	-	-	-
x_{126}	-	-	-	-	1	1	-	-	-	-
x_{128}	1	1	1	1	1	1	1	1	1	1
x_{132}	-	1	-	-	-	-	1	1	-	-
x_{135}	1	1	1	1	-	-	1	1	1	1
x_{142}	1	1	1	1	1	1	1	1	1	1
x_{143}	1	1	1	1	1	1	1	1	1	1
x_{145}	1	1	1	1	1	1	1	1	1	1
x_{144}	1	1	1	1	1	1	1	1	1	-

As mentioned before, the obtained selected features are ranked with respect to the number of occurrences of each feature in the 10 rounds and its criteria-or fitness-function. Table 5-7 shows the number of occurrences of each feature in the 10 iterations.

5.9 Classification

Various classification methods have been used for classifying mammogram masses into malignant or benign. In our experiment we use the Feed-Forward Artificial Neural Network (ANN) with back-propagation, which is one of the most popular techniques, as a classifier. More detail about ANN was introduced in section 3.1.5.

We use three levels to represent the forward neural network, the input layer with a number of neurons equal to the number of selected features, the output layer with a number of three nodes to represent the target classes –normal, benign and cancer-, and a hidden layer.

Table 5-7: The selected features and the number of times it is selected by each technique.

	Forward Sequential	Genetic Algorithm
x_{145}	10	10
x_{143}	9	10
x_{135}	7	8
x_{76}	5	5
x_{142}	5	10
x_{59}	4	10
x_{64}	2	10
x_{103}	1	2
x_{126}	1	2
x_{43}	1	10
x_{128}	1	10
x_{33}	1	10
x_{96}	1	10
x_{120}	2	2
x_{132}	0	3
x_{144}	0	9
x_{61}	0	10
x_{31}	0	10
x_{139}	3	0
x_{138}	2	0
x_{80}	2	0
x_{114}	1	0
x_{22}	1	0
x_{69}	1	0
x_{44}	1	0
x_{60}	1	0
x_{39}	1	0
x_{16}	1	0
x_{98}	1	0
x_{84}	1	0
x_{82}	1	0

In hidden layer, we faced the problem of what is the perfect number of nodes we have to use in the hidden layer? Too few nodes for hidden layer lead to under-fitting and too large number of nodes lead to over-fitting. There is no specific rule to determine the number of neurons -nodes- of hidden layer, so selecting the number of hidden layers neurons come down to trial and error. By trial and error we found that the best number of hidden layer neurons is the number of input + 2.

We use “Log-sigmoid” for transfer function, which is used to calculate the layer's output from its net input as follow:

$$\text{logsig}(n) = 1/(1 + \exp(-n)) \quad 5.2$$

Since we are using a feed-forward ANN, which is a gradient descent search, it will always suffer from local minima and the result will heavily depend on the initial values. We try to solve this problem and to get a better estimate of the performance by 10-fold cross validation.

We implement the feed-forward ANN four times, each time we use one of the four sets that are obtained in feature selection stage, which are SFF, GAF, USF, and ISF.

In the following we preview the classifications results using mentioned performance metrics:

1- Classification using SFF

The total number of the features selected by forward sequential technique is 27 features –as shown in Table 5-5 -; we use these features as input to the feed-forward ANN with 29 neurons in the hidden layer and three neurons at the output layer.

Training stage

Figure 5-7 shows the ROC curve for each of training, validation and test observations classifications. When we have a look on this figure we can judge that we have a very good classification with high performance as shown by the largest area under the ROC curve.

The Confusion matrix is shown in Figure 5-8, it shows that:

- 1- From 126 cancer cases, 105 are classified truly as cancer while the remained 21 are classified as benign,
- 2- From 105 benign cases, 89 are classified truly as benign, and 16 cases are classified in false as a cancer,
- 3- And 179 normal cases are classified truly as normal.

The overall classification accuracy is 91%, with 83.33% sensitivity and 84.7% specificity.

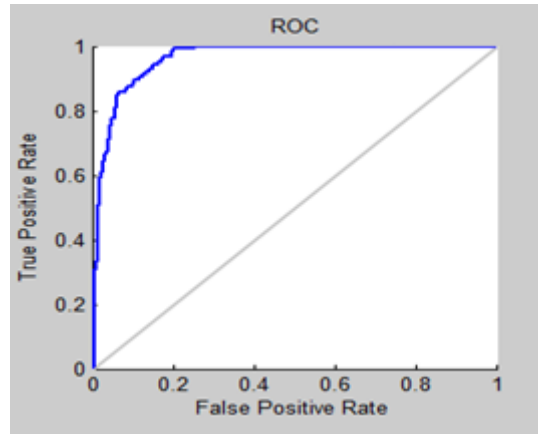


Figure 5-7: ROC curve analysis of FFANN classifier for training using SFF features.

Confusion Matrix				
Output Class	1	2	3	
	105 25.6%	16 3.9%	0 0.0%	86.8% 13.2%
	21 5.1%	89 21.7%	0 0.0%	80.9% 19.1%
	0 0.0%	0 0.0%	179 43.7%	100% 0.0%
				Target Class
				1 2 3
				83.3% 16.7% 84.8% 15.2% 100% 0.0% 91.0% 9.0%

Figure 5-8: Confusion matrix of FFANN classifier for training using SFF features.

Test stage

As mentioned before, we use two different and difficult datasets for testing the model obtained in training stage. The difficulties of the used test datasets lead to decrease the accuracy of classification. With dataset-1 the accuracy is 82% as shown in the confusion matrix in Figure 5-9, and with dataset-2 the accuracy is 86%.

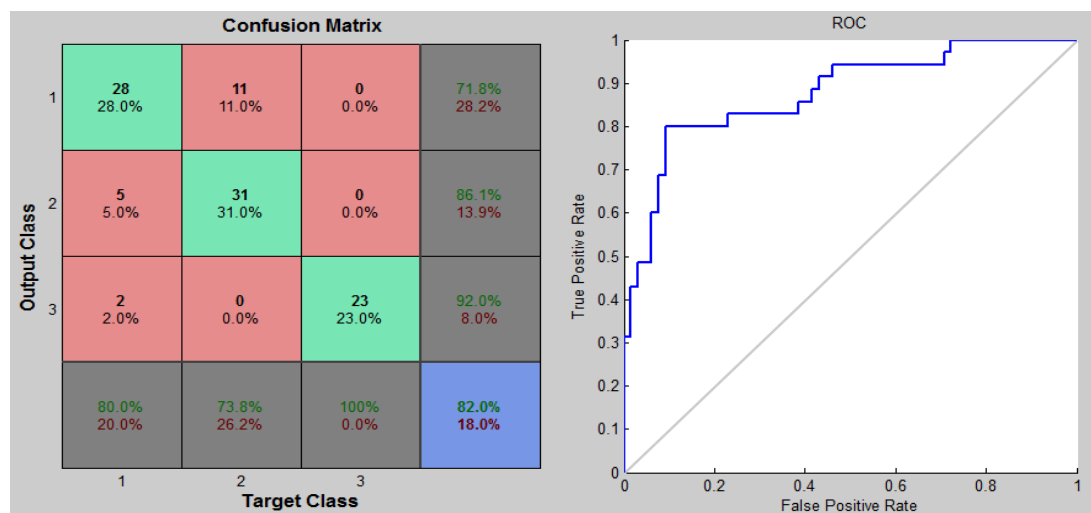


Figure 5-9: Confusion matrix and ROC curve of testing dataset-1 using SFF features.

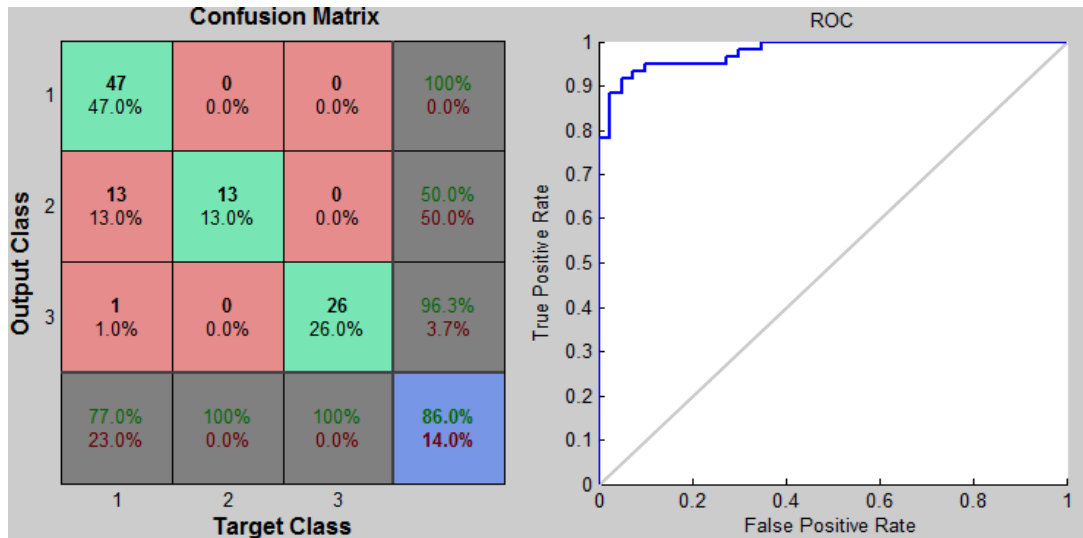


Figure 5-10: Confusion matrix and ROC curve of testing dataset-2 using SFF features.

2- Classification using GAF

Using Table 5-6, GA method selects 18 features as the best for the classification; we use these features as input to the feed-forward ANN with 20 neurons in the hidden layer and three neurons at the output layer.

Training stage

The classification performance is demonstrated by the ROC curve shown In Figure 5-11, as shown the area under ROC curve is large like that in Figure 5-7, which means that the performance here will be near that in case 1. Figure 5-12 shows the Confusion matrix, which can interpreted as follow:

- 1- From 126 cancer cases, 103 are classified truly as cancer while the remained 23 are classified as benign,
- 2- From 105 benign cases, 90 are classified truly as benign and 15 cases are classified in false as a cancer,
- 3- And 179 normal cases are classified truly as normal.

The overall classification accuracy is 90.7%, with 82% sensitivity and 85.7% specificity.

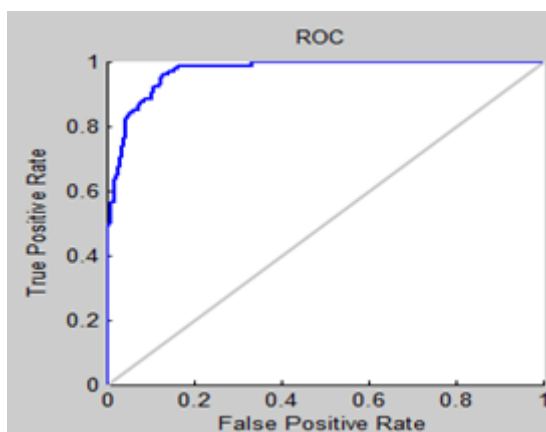


Figure 5-11: ROC curve analysis of FFANN classifier for training using GAF features.

Confusion Matrix				
Output Class	1	2	3	
	103 25.1%	15 3.7%	0 0.0%	87.3% 12.7%
	23 5.6%	90 22.0%	0 0.0%	79.6% 20.4%
	0 0.0%	0 0.0%	179 43.7%	100% 0.0%
				Target Class
				1 2 3

Figure 5-12: Confusion matrix of FFANN classifier for training using GAF features.

Testing stage

After training we test the created model using two different and difficult datasets. As shown in Figure 5-13, the accuracy with dataset-1 is 84%. In cancer cases, 31 of 35 cases are classified truly as cancer, 2 as benign and 2 as normal. In dataset-2, the accuracy is 86%, which is better than that in dataset-1, but both are less than the accuracy obtained with training stage. As mentioned before, the differences in accuracy returned to the difficulties of datasets. At the end of this Chapter we'll talk in more detail about these difficulties.

3- Classification using ISF

Shared features are the features that are selected by both sequential and GA techniques; the number of these features is 13 features. We create a feed-forward ANN with 13 neurons for input, 15 neurons for the hidden layer and 3 neurons for output layer.

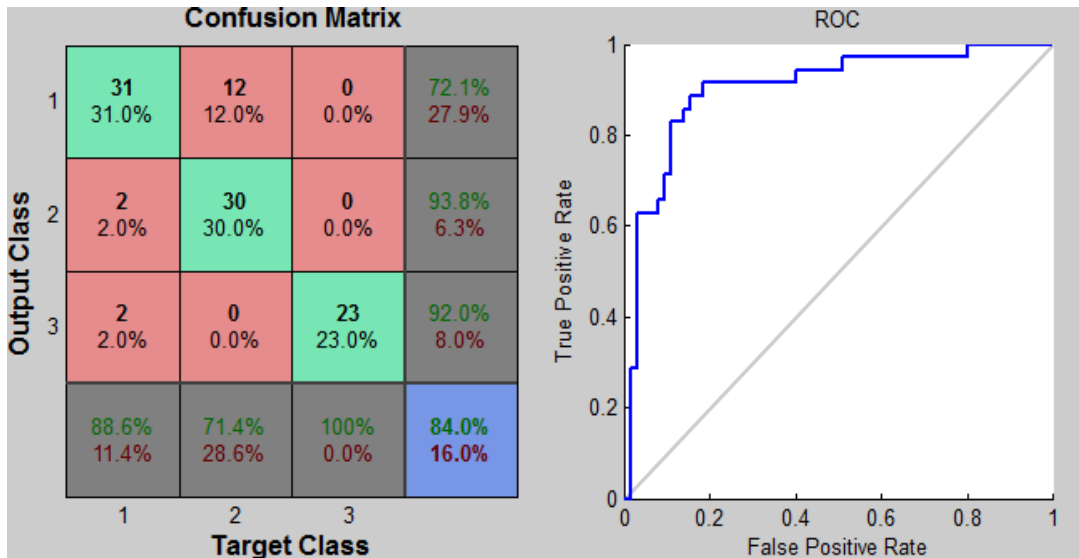


Figure 5-13: Confusion matrix and ROC curve of testing dataset-1 using GAF features.

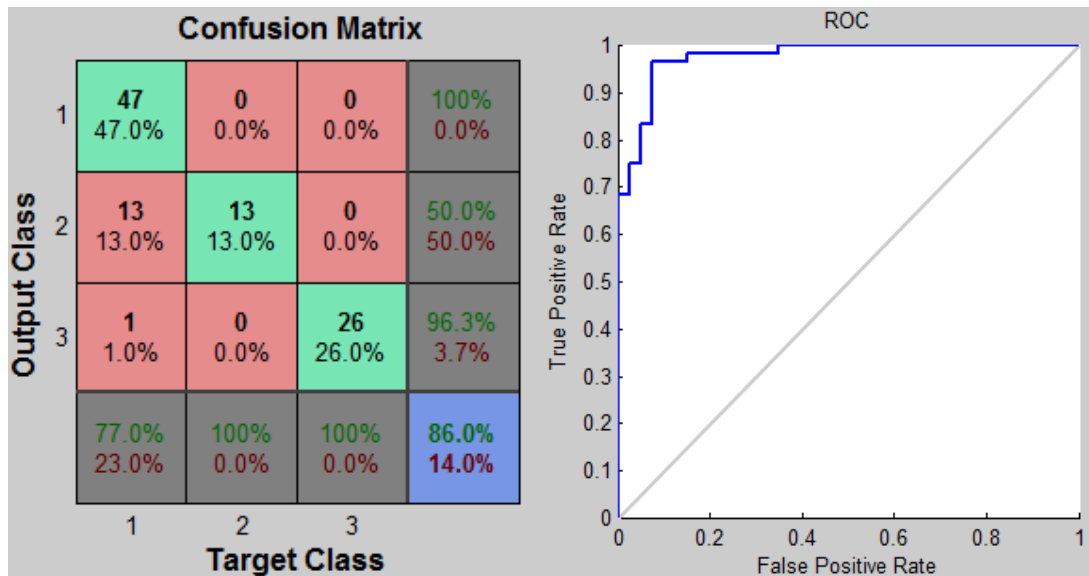


Figure 5-14: Confusion matrix and ROC curve of testing dataset-2 using GAF features.

Training stage

Figure 5-15 and Figure 5-16 shows the confusion matrix and the ROC curve for classification with ISF features:

- 1- 112 of 126 cancer cases, are classified truly as cancer while the remained 14 cases are classified as benign,
- 2- With 105 benign cases, 88 are classified truly as benign and 17 cases are classified in false as a cancer,
- 3- And 179 normal cases are classified truly as normal.

The overall classification accuracy is 92.4%, with 88.9% sensitivity and 83.81% specificity. As shown the results is better than that obtained with either GAF or

SFF. Figure 5-17 shows the time required to reach the best validation performance, which is just 14 epochs.

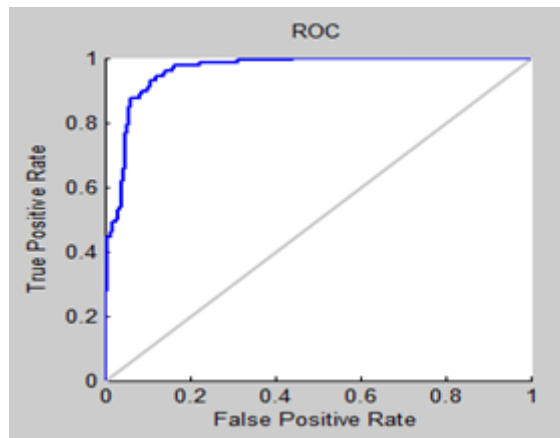


Figure 5-15: The ROC curve analysis of FFANN classifier for training using ISF features.

Test stage

As shown in Figure 5-18, the accuracy with dataset-1 is 86%. In cancer cases, 31 of 35 cases are classified truly as cancer, 2 as benign and 2 as normal. In dataset-2, the accuracy is 89% as demonstrated in Figure 5-19. In general the accuracy of classification with testing datasets is better with ISF than with either SFF or GAF. But it is less than the accuracy that we obtained with training dataset.

4- Classification using USF

We have 31 features as a total number of the selected features by each of sequential and GA techniques. In the following we show the experiment results with training and testing data set using USF.

Confusion Matrix				
Output Class	1	2	3	
	112 27.3%	17 4.1%	0 0.0%	86.8% 13.2%
	14 3.4%	88 21.5%	0 0.0%	86.3% 13.7%
	0 0.0%	0 0.0%	179 43.7%	100% 0.0%
				Target Class
				1
				2
				3
				92.4% 7.6%

Figure 5-16: Confusion matrix of FFANN classifier for training using ISF features.

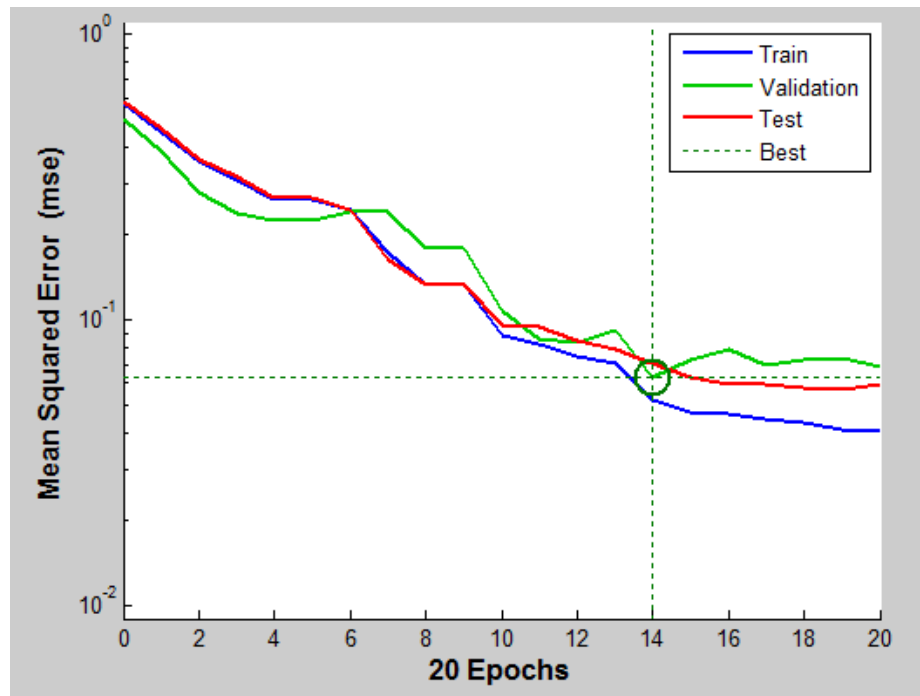


Figure 5-17: With shared features, the best validation performance is 0.062948 at epoch 14.

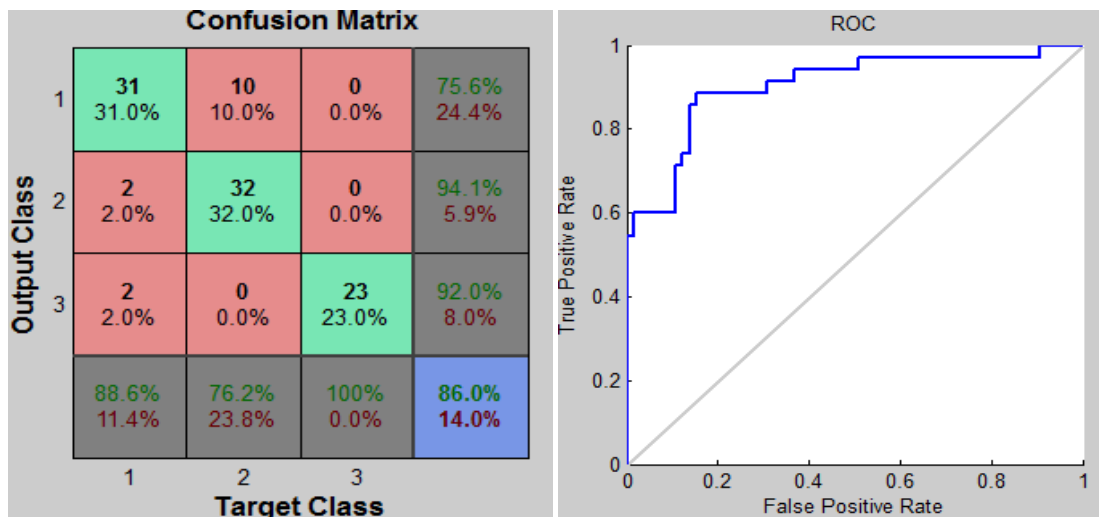


Figure 5-18: Confusion matrix and ROC curve of testing dataset-1 using ISF features.

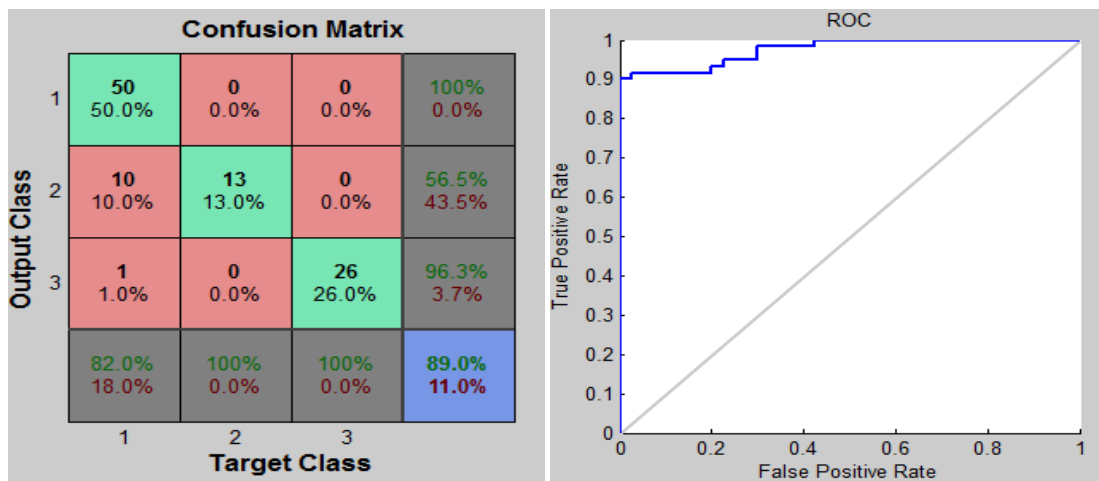


Figure 5-19: Confusion matrix and ROC curve of testing dataset-2 using ISF features.

Training stage

Figure 5-20 shows the ROC curve of the classification using USF, the area under the curve is so small which means that we have the best accuracy. The total accuracy reaches 96.3% as shown in the “all” confusion matrix in Figure 5-21; in which we have:

- 1- 117 of 126 cancer cases are classified truly as cancer while the remained 9 cases are classified as benign,
- 2- 99 of 105 benign cases are classified truly as benign while 6 cases are classified in false as a cancer,
- 3- And 179 normal cases are classified truly as normal.

So with union features we have the best performance, but it takes more time. The overall classification accuracy is 96.3%, with 92.9% sensitivity and 94.3% specificity. As shown in Figure 5-22, the classifier needs 40 epochs to reach the best validation performance.

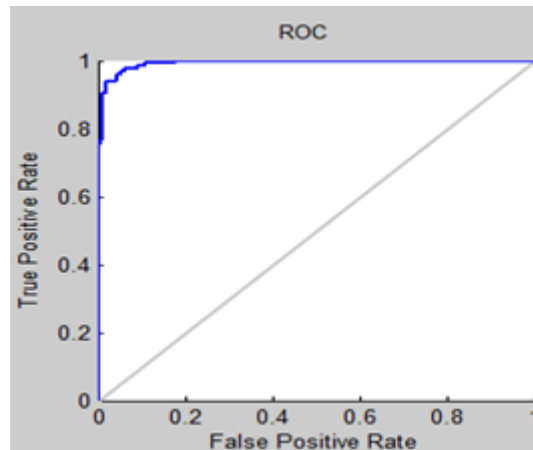


Figure 5-20: ROC curve analysis of FFANN classifier for training using USF features.

Test stage

As shown in Figure 5-23, the accuracy of the classification with dataset-1 is 88%. In cancer cases, 31 of 35 cases are classified truly as cancer, 2 as benign and 2 as normal. Figure 5-24 shows the confusion matrix and ROC curve for the classification with USF, The accuracy of the classification is 89%, it is better than that with SFF, GAF, or ISF.

5.10 Results Discussion

The best accuracy obtained with USF features; with training dataset the accuracy is 96%, but this accuracy decreased to 89% with test datasets. As mentioned before, the main reason of this difference is the difficulties with testing datasets. These difficulties could be described and demonstrated by Figure 5-25, which shows four misclassified images described as follow:

Output Class	1	2	3	
	117 28.5%	6 1.5%	0 0.0%	95.1% 4.9%
	9 2.2%	99 24.1%	0 0.0%	91.7% 8.3%
	0 0.0%	0 0.0%	179 43.7%	100% 0.0%
	92.9% 7.1%	94.3% 5.7%	100% 0.0%	96.3% 3.7%
				Target Class
	1	2	3	

Figure 5-21: Confusion matrix for FF-ANN classifier using USF features.

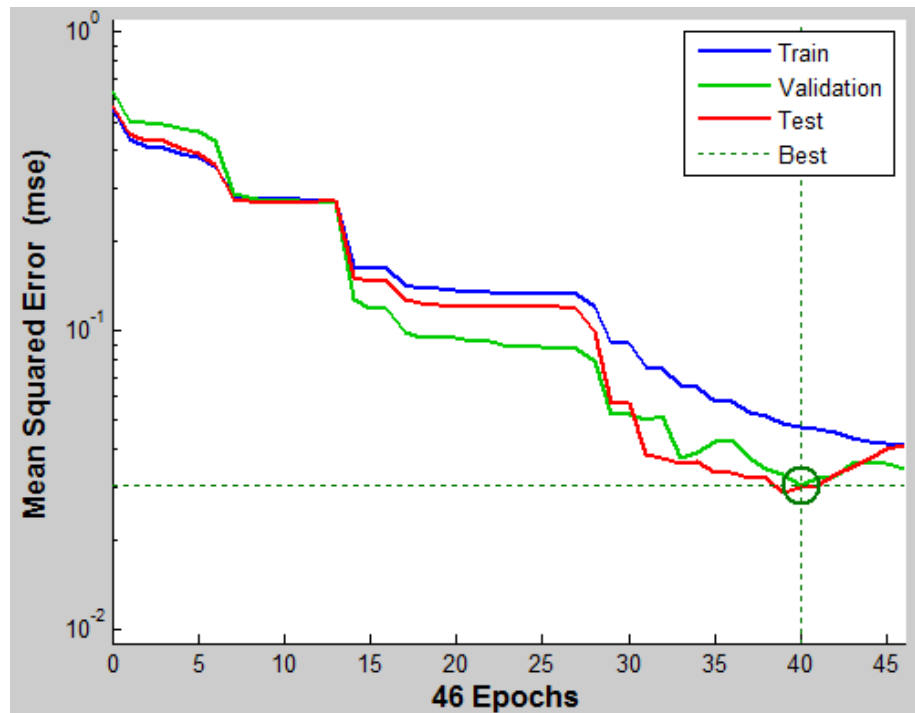


Figure 5-22: With USF features, the best validation performance is 0.030551 at epoch 40.

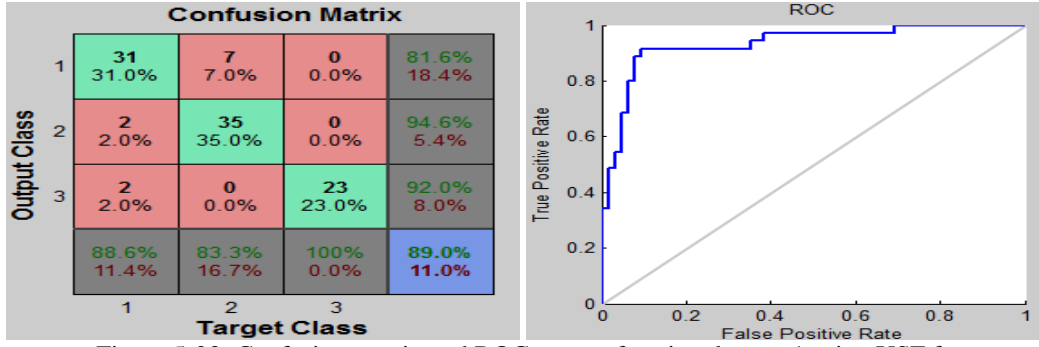


Figure 5-23: Confusion matrix and ROC curve of testing dataset-1 using USF features.

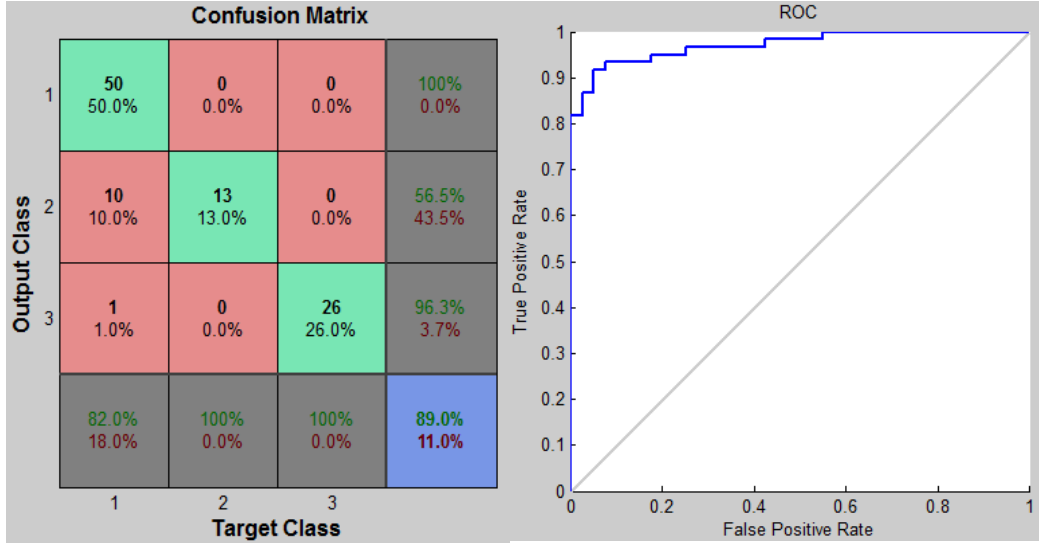


Figure 5-24: Confusion matrix and ROC curve of testing dataset-2 using USF features.

- a- Image (a) which represents the case “C0125-Right-MLO” is a cancer case in its early stage with an oval shape and ill-defined margins. This image is classified in false as benign.
- b- Image (b) represents the case “C0085-Right-CC” is a cancer case with irregular shape and spiculated margins. This image is classified in false as benign.
- c- Image (c) is a cancer case represents “A1144-Right-MLO”, it has an irregular shape and ill-defined margins, and is classified as benign.
- d- C0029LeftMLO is a benign case with oval shape and ill-defined margins.

Images (a), (c) and (d) has ill-defined margins, in which we have unrestrained proliferation of cells in irregular directional pattern. In early stages it is so difficult to differentiate between benign and cancer and also it is so difficult to segment the mass in perfect. In our thesis, 14% of training dataset -56 cases- has ill-defined margins, 49 of them are classified truly. But in testing datasets, we have 50% of the cases with ill-defined margins; most of them are in early stage.

The difficulties of image (b) is in its spiculated margins, which make the segmentation task so difficult, and sure the classification will be influenced by the accuracy of segmentation results.

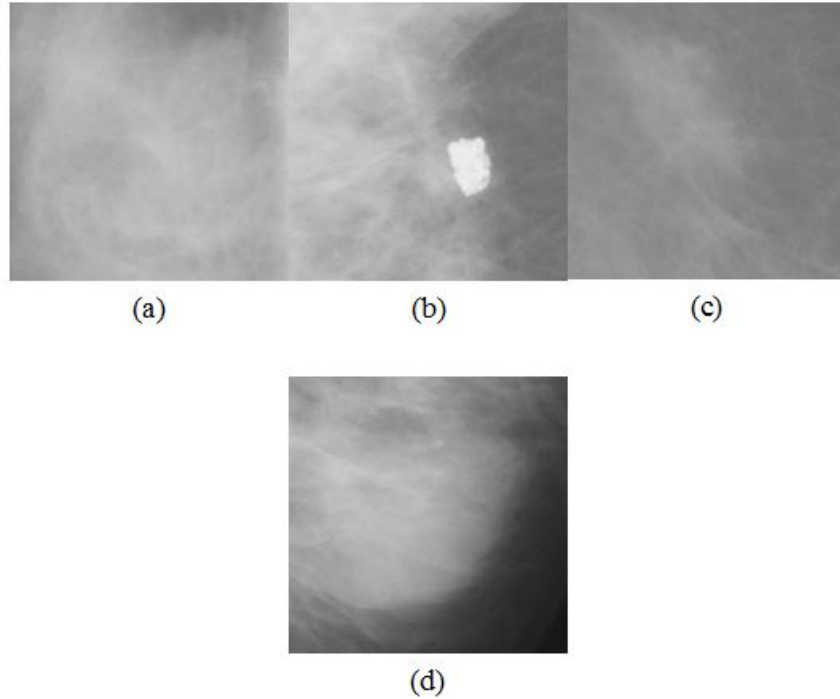


Figure 5-25: Examples of misclassified mammogram images

In Table 5-8, we show a simple comparison in accuracy between our method and others. Actually, direct comparison of these systems is so difficult because most of these studies done on different databases and different cases.

Table 5-8: A simple comparison in accuracy between our method and others.

Author	Training dataset		Test datasets			
	size	accuracy	size	accuracy	size	accuracy
Pohlman [16]	51	76%-93%	-	-	-	-
Hadjiiski [34]	348	81%				
Jiang [15]	211	72.7%	-	-	-	-
Surendiran [27]	300	87%	-	-	-	-
Zheng [11]	3000	87%				
Proposed method	410	96%	100	89%	100	89%

Chapter 6

Conclusion and Future Work

6.1 Summary and Concluding Remarks

Breast cancer is the most common form of cancer among women and is the second leading of death after lung cancer. Early detection and diagnosis of breast cancer increase the treatment options and a cure is more likely.

One of the most effective tools for early detection of breast cancer is the mammography, but Visual interpretation of a mammogram is so difficult and the probability of false negatives is high. Two thirds of these false negative results are due to missed lesions that are evident retrospectively. Also, a significant level of false positives was reported, which means a high proportion of biopsies are performed on benign lesions.

Avoiding benign biopsies would spare women anxiety, discomfort, and expense. So it became very important to find new method aids in detection and diagnosis mammogram suspicious regions. The computer-aided diagnosis (CADx) is a helpful system used to classify the detected regions into malignant or benign categories to help the radiologists in recognizing the next step, biopsy or short-term follow-up mammography.

The performances of different classifying algorithms can come to the same level in detecting mammogram lesions with the same features from the same mammographic database. So the performance of CADx depends more on the optimization of the feature extraction and selection than the classification methods. To obtain this, we concentrate on feature extraction and selection.

We use the Digital Database for Screening Mammography (DDSM) as a source of the mammographic images, which we use in our experiments.

We do some enhancement on pre-existing feature extraction methods, which are GLCDM and RDM, and we propose a new feature extraction method. We call it: Square Centroid Lines Gray level distribution Methods (SCLGM). By choosing four

centroid lines we compute six mathematical measures, which we use to extract 79 features.

After feature extraction, it is important to apply feature selection to avoid the curse of dimensionality and to reduce the complexity of the CADx system by eliminating irrelevant features and selecting the best discriminative features.

It is very difficult to predict which feature or feature combinations will achieve better in classification rate. We will have different performances as a result of different feature combinations. In addition, using excessive features may degrade the performance of the algorithm and increase the complexity of the classifier. Therefore, we have to select an optimized subset of features from a large number of available features.

In our thesis we choose two feature selection methods: the first is categorized into the sequential search methods, which is the forward sequential feature selection technique, and the second method is categorized into the randomized search method, which is the Genetic feature selection Algorithm. These two methods select 31 features, 18 of the selected features are from our proposed feature extraction method (SCLGM).

In classification, we repeat the classification experiment four times, in each time we change the input of selected features to be one of the following: features selected by sequential technique, features selected by GA technique, features selected by both sequential and GA techniques and the total features.

The performance of the classification with features selected by GA near to the performance obtained with features selected by forward selection method. But with shared features, we get better performance and in less time; we reach the best performance.

The best performance of all is obtained when we use all of the selected features by both sequential and GA techniques as input to the classifier. But regarding the time required to reach the best performance, it will be larger than that if we use little features.

6.2 Future Work

The performance could be enhanced more and more by extracting and selecting the best and the most discriminative features so for future work we suggest the following:

- Feature extraction performance is greatly affected by the segmentation process. Here in our thesis we use gray level region grow methods for segmentation, and it does well, but not with ill-defined cases. We can obtain better results if we do enhancement on segmentation methods, so enhancing segmentation methods will be one of our future works.
- Enhancing our new proposed feature extraction technique SCLGM by finding more measures to extract more features will be one of our future works.
- In the experiments, in feature extraction using our new technique, we choose just four central lines to compute the measures and then to extract 79 feature, but if we chooses 8 central lines or more the extracted features will be more discriminative.
- We apply SCLGM on mammograms with mass types. In the future work we will apply it on the calcification cancer type, and it will extract a discriminative feature and will do best.
- After feature extraction we have to enhance the feature selection method, as shown in our experiments the selected features play the main role in enhancing the classifications performance.
- Finally, we should do comparison between different classifiers' methods.

Bibliography

- [1] (2011) Cancer facts and figures 2011. [Online].
<http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancer-key-statistics>
- [2] MOH, "Cancer incidence in the Gaza Strip 1998 - 2008. ," Ministry of Health, Gaza, Palestine, 2009.
- [3] American College of Radiology, "Breast imaging reporting and data system:BI-RADS atlas," American College of Radiology, Reston, atlas 2003.
- [4] Sergios Theodoridis and Konstantinos Koutroumbas, *Pattern Recognition*, 4th ed., Elsevier's Science & Technology, Ed. California, United States of America: Academic Press , 2009.
- [5] M. L. Giger, N. Karssemeijer, and S. G. Armato, "Computer-aided diagnosis in medical imaging," *IEEE*, pp. 1205–1208, 2001.
- [6] C. J. Vyborny, M. L. Giger, and R. M. Nishikawa, "Computer-aided detection and diagnosis of breast cancer," *Radiol*, pp. 725–740, 2000.
- [7] B. Zheng, Y. H. Chang, X. H. Wang, and W. F. Good, "Comparison of artificial neural network and Bayesian belief network in a computer assisted diagnosis scheme for mammography," in *IEEE International Conference on Neural Networks*, 1999, pp. 4181–4185.
- [8] M.P. Sampat, M.K. Markey, and A.C. Bovik, *Computer-Aided Detection and Diagnosis in Mammography*, A.C. Bovik, Ed. Amsterdam: Elsevier Academic Press, 2005.
- [9] T. S. Subashini, V. Ramalingam, and S. Palanivel, "Automated assessment of breast tissue density in digital mammograms," *Elsevier*, pp. 33-43, 2010.
- [10] J. Marti, J. Freixenet, X. Mu noz, and A. Oliver, "Active region segmentation of mammographic masses based on texture, contour and shape features.," in *Springer-Verlag*, Berlin Heidelberg, 2003, pp. 478-485.
- [11] Bin Zheng et al., "Interactive Computer-Aided Diagnosis of Breast Masses: Computerized Selection of Visually Similar Image Sets From a Reference Library.," *Acad Radio*, vol. 14, pp. 917–927, 2007.
- [12] JL Viton, M. Rasigni, G. Rasigni, and A. Liebaria, "Method for characterizing masses in digital mammograms," *Opt. Eng.*, vol. 35, no. 12, pp. 3453–3459, 1996.
- [13] L. H. Li, W. Qian, and L. P. Clarke, "Digital mammography: computer-assisted diagnosis method for mass detection with multiorientation and multiresolution wavelet transforms," *Acad. Radiol*, no. 4, pp. 724–731, 1997.

- [14] L. S. Reddy, R. Reddy, CH. Madhu, and C. Nagaraju, "a novel image segmentation technique for detection of breast cancer," *International Journal of Information Technology and Knowledge Management*, vol. 2, no. 2, pp. 201-204, 2010.
- [15] Luan Jiang, Enmin Song, Xiangyang Xu, Guangzhi Ma, and Bin Zheng, "Automated Detection of Breast Mass Spiculation Levels and Evaluation of Scheme Performance," *Acad Radiol*, vol. 15, pp. 1534–1544, 2008.
- [16] S. Pohlman, K. A. Powell, N. A. Obuchowski, W. A. Chilcote, and S. Grundfest-Bronlatowski, "Quantitative classification of breast tumors in digitized mammograms," *Medical Physics*, vol. 23, pp. 1337-1345, 1996.
- [17] K. Bovis and S. Singh, "Detection of masses in mammograms using texture features.," in *15th International Conference on Pattern Recognition (ICPR'00)*, 2000, pp. 2:2267.
- [18] Indra Kanta Maitra, Sanjay Nag, and Samir Kumar Bandyopadhyay, "Identification of Abnormal Masses in Digital Mammography Images," *International Journal of Computer Graphics*, vol. 3, no. 1, pp. 17-29, May 2011.
- [19] L. Blot and R. Zwiggelaar, "Extracting background texture in mammographic images: a co-occurrence matrices based approach.," in *Proceedings of the 5th International Workshop on Digital Mammography*, Toronto(Canada), 2000, pp. 142-148.
- [20] WEI-MING CHEN et al., "3-D ULTRASOUND TEXTURE CLASSIFICATION USING RUN DIFFERENCE MATRIX," *World Federation for Ultrasound in Medicine & Biology*, vol. 31, no. 6, pp. 763–770, 2005.
- [21] K. Aoki and M. Kudo, "Feature and Classifier Selection in Class Decision Trees," *SSPR&SPR*, pp. 562–571, 2008.
- [22] M. L. Giger, Z. Huo, M. A. Kupinski, and C. J. Vyborny, *Computer-aided diagnosis in mammography*, J.M. Fitzpatrick M. Sonka, Ed.: SPIE Press, 2000, vol. 2.
- [23] H. D. Cheng et al., "Approaches for Automated Detection and Classification of Masses in Mammograms," *Pattern Recognition*, vol. 39, no. 4, pp. 646-668, 2006.
- [24] F. N. Che, M. C. Fairhurst, C. P. Wells, and M. Hanson, "Evaluation of a two-stage model for detection of abnormalities in digital mammograms," in *Proceedings of the 1996 IEE Colloquium on Digital Mammography*, London, UK, 1996.
- [25] B. Sahiner et al., "Image feature selection by a genetic algorithm: application to classification of mass and normal breast tissue on mammograms," *Med. Phys.*, no. 23, pp. 1671–1684, 1996.
- [26] B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Design of a high-sensitivity classifier based on a genetic algorithm: application to computer-aided

- diagnosis," *Phys. Med. Biol.*, pp. 2853–2871, 1998.
- [27] B. Surendiran and A. Vadivel, "Feature Selection using Stepwise ANOVA Discriminant Analysis for Mammogram Mass Classification," *International J. of Recent Trends in Engineering and Technology*, vol. 3, no. 2, pp. 55-57, 2010.
 - [28] D. Wei et al., "Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis," *Phys. Med.*, no. 22, pp. 1501–1513, 1995.
 - [29] M. A. Kupinski and M. L. Giger, "Feature selection and classifiers for the computerized detection of mass lesions in digital mammography," in *IEEE International Conference on Neural Networks*, 1997, pp. 2460–2463.
 - [30] W.J. H. Veldkamp, N. Karssemeijer, and J.D. M. Otten, "Automated classification of clustered microcalcifications into malignant and benign types," *Med. Physics*, no. 21, pp. 2600–2608, 2000.
 - [31] O. Tsujii, M. T. Freedman, and S. K. Mun, "Classification of microcalcifications in digital mammograms using trend-oriented radial basis function neural network," *Pattern Recognition*, no. 32, pp. 891–903, 1999.
 - [32] M. Kallergi, "Computer-aided diagnosis of mammographic microcalcification clusters," *Med. Physics*, no. 31, pp. 314–326, 2004.
 - [33] M. VASANTHA, DR. V. SUBBIAH BHARATHI, and R. DHAMODHARAN, "Medical Image Feature, Extraction, Selection And Classification," *International Journal of Engineering Science and Technology*, vol. 2, no. 6, pp. 2071-2076, 2010.
 - [34] L. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, and M. Helvie, "Classification of malignant and benign masses based on hybrid ART2LDA approach," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 1178-1187, 1999.
 - [35] Shinsuke Katsuhara, Hitoshi Futamura, Satoshi Kasai, Takako Morita, and Tokiko Endo3, "Computerized Scheme for Focal Asymmetric Densities on Mammograms by Use of Geometric and Texture Analyses," in *Springer-Verlag Berlin Heidelberg 2008*, Berlin, 2008, pp. 315–322.
 - [36] Brijesh Verma, Peter McLeod, and Alan Klevansky, "A novel soft cluster neural network for the classification of suspicious areas in digital mammograms," *Pattern Recognition*, no. 42, pp. 1845 -- 1852, 2009.
 - [37] M. Karnan and K. Thangavel, "Computer Aided Diagnosis in Digital Mammograms: Detection of Microcalcifications by Meta Heuristic Algorithms ," *GVIJ Journal*, vol. 5, no. 7, pp. 41-55, 2005.
 - [38] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, 2nd ed., Inc.

- Prentice-Hall, Ed. New Jersey, United States of America: Tom Robbins, 2002.
- [39] H.D. Cheng et al., "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognition Society*, no. 39, pp. 646 – 668, 2006.
- [40] P.E. Hart, and D.G. Stork R.O. Duda, *Pattern Classification*. New York, United States of America: John Wiley & Sons, 2001.
- [41] Wikipedia. [Online]. http://en.wikipedia.org/wiki/Support_vector_machine
- [42] Antonin Slaby, "ROC Analysis with Matlab," in *Proceedings of ITI 2007 29th Int. Conf. on Information Technology Interfaces*, Czech republic, 2007, pp. 191-196.
- [43] B. Julesz, "Experiments in the Visual Perception of Texture," *Scientific American*, vol. 4, no. 232, pp. 34-43, 1975.
- [44] Aristofanes C. Silva, Anselmo C. Paiva, Paulo C.P. Carvalho, and Marcelo Gattass, "Semivariogram and SGLDM Methods Comparison for the Diagnosis of Solitary Lung Nodule," *Springer*, pp. 479–486, 2005.
- [45] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image," *SMC* 3, pp. 610–621, 1973.
- [46] University of South Florida. (2008) Digital Database for Screening Mammography (DDSM). [Online]. <http://marathon.csee.usf.edu/Mammography/Database.html>
- [47] Hanchuan Peng, Fuhui Long, and Chris Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 27, no. 8, pp. 1226-1238, 2005.
- [48] Tony F. Chan and Luminita A. Vese, "Active Contours Without Edges," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 10, no. 2, pp. 266-277, 2001.
- [49] B. Zheng, Y. H. Chang, X. H. Wang, and W. F. Good, "Comparison of artificial neural network and Bayesian belief network in a computer-assisted diagnosis scheme for mammography," in *IEEE International Conference on Neural Networks*, 1999, pp. 4181–4185.
- [50] M. Hajek, M. Dezortova, A. Materka, and R. Lerski, *texture analysis for magnetic resonance imaging*. Prague: HartPress, 2006.