

**MACHINE LEARNING-BASED CLASSIFICATION OF
BREAST DENSITIES**

BY

Abdullah Hussein Abdullah Owaidh

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE


In

COMPUTER SCIENCE


November 2014

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN- 31261, SAUDI ARABIA
DEANSHIP OF GRADUATE STUDIES

This thesis, written by **Abdullah Hussein Owaidh** under the direction his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN COMPUTER SCIENCE**.



Dr. Adel F. Ahmed
Department Chairman




Dr. Salam A. Zummo
Dean of Graduate Studies

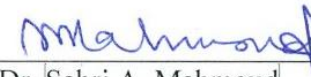
19/12/14

Date






Dr. Lahouari Ghouti
(Advisor)



Dr. Sabri A. Mahmoud
(Member)



Dr. Wasfi G. Al-Khatib
(Member)

© Abdullah Hussein Abdullah Owaidh
2014

Dedication

*To My Parents,
My Wife and Lovely Sons Mohammed & Hussein*

ACKNOWLEDGMENTS

First of all, the praise and thanks be to ALLAH, the Almighty, for his uncountable favours. I would like to thank King Fahd University for Petroleum and Minerals (KFUPM), and Hadramout Establishment for Human Development for giving me the opportunity to complete my study in KFUPM University.

I am pleased to express my deepest thanks to my thesis advisor, Dr. Lahouari Ghouti, for all his guidance, support and advice without which this work would not have been possible. I would like also to thank the committee members for their invaluable comments and support.

Thanks are also due to Galal Bin Makhasen for his help and comments. I need to thank all individuals who contributed and assisted in the completion of this thesis.

A special thanks goes to Sheik Eng. Abdullah Ahmed Bugshan for his encouragement and support to complete my study.

Finally, I must thank my parents, without their faith and encouragement I would not be able to succeed in my life.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	VII
TABLE OF CONTENTS.....	VIII
LIST OF TABLES	XI
LIST OF FIGURES	XII
LIST OF ABBREVIATIONS.....	XIV
ABSTRACT.....	XVI
ملخص الرسالة.....	XVII
CHAPTER 1 INTRODUCTION	1
1.1 Survival Rate	3
1.2 Breast Anatomy	3
1.3 Breast Density.....	4
1.4 Breast Imaging.....	6
1.4.1 Mammography Screening	6
1.4.2 Ultrasound Imaging.....	9
1.4.3 Magnetic Resonance Imaging (MRI).....	10
1.5 Breast Imaging Reporting and Database System (BI-RADS)	11
1.6 Computer-Aided Diagnosis (CAD)	13
1.7 Classification Performance Measures	14
1.8 Problem Statement.....	16
1.9 Thesis outline.....	16
CHAPTER 2 LITERATURE REVIEW	18

2.1	Quantitative Approaches.....	18
2.2	Qualitative Approaches.....	19
2.2.1	Matrix Factorization Methods.....	20
2.2.2	Global Histogram Methods	21
2.2.3	Texture Analysis Methods	26
2.2.4	Summary of the literature	31
CHAPTER 3 PROPOSED SOLUTION.....		34
3.1	Methodology.....	34
3.1.1	Preprocessing and Segmentation Step.....	36
3.1.2	Feature Extraction	40
3.1.3	Classification.....	54
CHAPTER 4 EXPERIMENT RESULTS.....		60
4.1	Databases and Tools	60
4.1.1	Mammographic Image Analysis Society (MIAS) Database	60
4.1.2	Development Tools	61
4.1.3	Classification.....	62
4.1.4	Validation and Classification Performance	62
4.2	Results	63
4.2.1	PCA Results	64
4.2.2	2D PCA Results	69
4.2.3	SVD Results.....	75
4.2.4	NMF Results	80
4.2.5	Mean Density Results	84
4.2.6	LBP Results	87
4.3	Conclusion.....	89

CHAPTER 5 CONCLUSION AND FUTURE WORK	93
5.1 Future Directions	94
REFERENCES	95
VITAE.....	100

LIST OF TABLES

Table 1-1: 5-Year survival rate	3
Table 1-2: BI-RADS tissue density classes.	12
Table 1-3: Assessment categories of BI-RADS.....	13
Table 1-4: Confusion matrix in binary classification.	15
Table 2-1: MIAS-based statistical features	22
Table 2-2: Summary of literature review.....	31
Table 3-1: PSNR values for different sizes of MIAS mammograms.	39
Table 3-2: PSNR values for different sizes of MIAS patches.	39
Table 4-1: Estimated time for feature extraction using the PCA method.	69
Table 4-2: Estimated time for 2D PCA extraction.....	74
Table 4-3: Estimated time for SVD extraction.	79
Table 4-4: Estimated time for feature extraction using the NMF factorization.....	83
Table 4-5: SVM performance measures using full mammograms and the Mean Density method.....	85
Table 4-6: KNN performance measures using full mammograms and the Mean Density method.....	86
Table 4-7: SVM performance measures using mammogram patches and the Mean Density method. ...	86
Table 4-8: KNN performance measures using mammogram patches and the Mean Density method. ...	86
Table 4-9: Estimated time for feature extraction using the Mean Density method.	87
Table 4-10: SVM performance measures using full mammograms and the NMF+LBP method.	87
Table 4-11: KNN performance measures using full mammograms and the NMF+LBP method.	88
Table 4-12: SVM performance measures using mammogram patches and the MF+LBP method.	88
Table 4-13: KNN performance measures using mammogram patches and the MF+LBP method.	89
Table 4-14: Estimated time for feature extraction using the NMF+LBP method.	89
Table 4-15: Benchmarking of the proposed approach.....	90
Table 4-16: Mumber of basis used in each method.	92

LIST OF FIGURES

Figure 1-1: Breast structure	4
Figure 1-2: A DCIS visible in a fatty dense mammography	5
Figure 1-3: Micro-calcification detected on mammogram	8
Figure 1-4: Macro-calcification detected on mammogram	8
Figure 1-5: CC view (left) MLO view (right) in mammogram screening.	9
Figure 1-6: A tissue shown as cancer in ultrasound imaging (false-positive)	10
Figure 1-7: (A) Dense breast in mammogram (B) MRI image.	11
Figure 1-8: Mammograms examples of BI-RADS classes from TYPE I (left) to TYPE IV (right).	12
Figure 1-9: Image processing typical steps in CAD systems.	14
Figure 2-1: MR8 filter bank.	27
Figure 2-2: Textons Learning stage	28
Figure 2-3: Texton Modeling Stage	29
Figure 3-1: Proposed breast density classification system.	35
Figure 3-2: Extracting pectoral muscle and labels from a mammogram.	36
Figure 3-3: A 300x300 pixels mammographic patch.	37
Figure 3-4: PCA Dimension reduction	42
Figure 3-5: Mean image of mammograms.	44
Figure 3-6: Top 70 PCA bases.	44
Figure 3-7: NMF bases.	49
Figure 3-8: Histogram of three densities, (Left) fatty glandular, (Middle) fatty, (Right) dense.	53
Figure 3-9 : Calculating LBP for pixel	54
Figure 3-10: Linear SVM structure.	56
Figure 3-11: Classification using KNN (K=3).	59
Figure 4-1: Cumulative sum of eigenvalues used in PCA decomposition.	64
Figure 4-2: Cumulative sum of eigenvalues of mammogram patches used in PCA decomposition.	65
Figure 4-3 PCA results on mammograms using SVM	66
Figure 4-4: PCA results using full mammograms and the KNN method.	66
Figure 4-5: PCA results using mammogram patches and the SVM method.	67
Figure 4-6: PCA results using mammogram patches and the KNN method.	68
Figure 4-7: Cumulative sum of eigenvalues on mammograms used in 2DPCA decomposition.	70

Figure 4-8: Cumulative sum of eigenvalues on patches used in 2DPCA decomposition.	71
Figure 4-9: 2DPCA results using full mammograms and the SVM method.	72
Figure 4-10: 2DPCA results using full mammograms and the KNN method.	72
Figure 4-11: 2DPCA results using mammogram patches and the SVM method.	73
Figure 4-12: 2DPCA results using mammogram patches and the KNN method.	74
Figure 4-13: Cumulative sum of eigenvalues in full mammogram factorization.....	75
Figure 4-14: Cumulative sum of eigenvalues in mammogram patches factorization.	76
Figure 4-15: SVD results using full mammograms and the SVM method.....	76
Figure 4-16: SVD results using full mammograms and the KNN method.....	77
Figure 4-17: SVD results using mammogram patches and the SVM method.....	78
Figure 4-18: SVD results using mammogram patches and the KNN method.....	79
Figure 4-19: NMF results using full mammograms and the SVM method.	81
Figure 4-20: NMF results using full mammograms and the KNN method.	81
Figure 4-21: NMF results using mammogram patches and the SVM method.	82
Figure 4-22: NMF results using mammogram patches and the KNN method.	83
Figure 4-23: A segmented fatty mammoram.	84
Figure 4-24: A segmented dense mammogram.	85
Figure 4-25: Misclassified brighter fatty mammogram.	91
Figure 4-26: Misclassified darker dense mammogram.....	91

LIST OF ABBREVIATIONS

2DPCA	:	2 Dimensional Principle Component Analysis
AUC	:	Area under Curve
BIF	:	Basic Image Feature
BI-RADS	:	Breast Imaging Reporting and Database System
CAD	:	Computer Aided Diagnosis
CBIR	:	Content-Based Image Retrieval system
CC	:	Cranio-Caudal view
DCIS	:	Ductal Carcinoma in Situ
DNA	:	Deoxyribonucleic Acid
FN	:	False Negative
FP	:	False Positive
FPR	:	False Positive Rate
KNN	:	K-Nearest Neighbor
LBP	:	Local Binary Pattern
LCIS	:	Lobular Carcinoma in Situ
LDA	:	Linear Discriminant Analysis
LGA	:	Local Grey level Appearances
MD	:	Mammography Density
MIAS	:	Mammographic Image Analysis Society
MLO	:	Medio-Lateral Oblique view
MR8	:	Maximum Response 8 filter

MRI	:	Magnetic Resonance Imaging
MSE	:	Mean Square Error
NMF	:	Non-negative Matrix Factorization
PCA	:	Principle Component Analysis
PSNR	:	Peak Signal-to-Noise Ratio
RBF	:	Radial Basis Function
ROC	:	Receiver Operating Characteristics
ROI	:	Region of Interest
SIFT	:	Scale Invariant Feature Transform
SVD	:	Singular Value Decomposition
SVM	:	Support Vector Machine
TN	:	True Negative
TP	:	True Positive
TPR	:	True Positive Rate
TSDM	:	Texton Spatial Dependence Matrix
US	:	Ultrasound Imaging

|

ABSTRACT

Full Name : Abdullah Hussein Abdullah Owaith
Thesis Title : Machine Learning Based Classification of Breast Densities
Major Field : Computer Science
Date of Degree : [November 2014]

Mammographic breast density describes the amount of fibro glandular tissue in the breast. Dense breast has more tissue than fat. The breast density is one of the strongest indicators of the increasing risk of developing breast cancer. Higher density breasts also decrease the sensitivity of mammography screening due to the tissue masking effect.

However, visual inspection of mammograms is recognized to be subjective and varies from one radiologist to another. Several research studies have been conducted to automate the breast density classification.

Breast Imaging Reporting and Data System (BI-RADS[®]) is a standard classification system for mammography density reporting. It is developed by the American College of Radiology (ACR). BI-RADS provides four categories for breast densities based on the visual assessment by radiologists.

In this work, a successful breast density classification system is designed and developed to classify mammographic breast density into two categories- fatty and dense. The proposed system uses: Principle Component Analysis (PCA), 2D-PCA, Singular Value Decomposition (SVD), Nonnegative Matrix factorization (NMF), Threshold to extract features. Then, these features are thresholds for classification purpose. Support Vector machine (SVM) and K-Nearest Neighbour (KNN) techniques are used in the classification stage. The results of our system are encouraging, and pave the way for a new approach for breast density classification. |

ملخص الرسالة

الاسم الكامل: عبدالله حسين عبدالله عويض

عنوان الرسالة: تصنيف كثافة الثدي اعتماداً على تعلم الآلة

التخصص: علوم حاسوب

تاريخ الدرجة العلمية:

يصف تصنيف كثافة الثدي الاشعاعي كمية الانسجة الموجودة في الثدي. لذا يعتبر الثدي ذا كثافة عالية عندما تكون كمية الانسجة اكثر من الدهون. حيث ان كثافة الثدي تعتبر من العوامل المهمة التي تشير الى احتمالية الاصابة بالسرطان . كما وان هذه الكثافة تعتبر ايضاً عائقاً للكشف عن السرطان عند الفحص بالاشعة لكونها تحجب الاشعة وتمنع ظهور السرطان فيها.

مع ذلك، اعتماد التشخيص البصري يعتبر اكثر موضوعية وقد تختلف النتائج من شخص لآخر، لهذا عمدت كثير من الدراسات والابحاث الى اتمتت هذه العملية وتصنيف كثافة الثدي تلقائياً.

قامت الكلية الامريكية للطب الاشعاعي بتطوير نظاماً قياسياً لتصنيف كثافة الثدي يسمى نظام تصوير الثدي التقارير والبيانات BI-RADS يقوم على تصنيف الكثافات الى اربعة مجموعات معتمداً في ذلك على التشخيص البصري لطبيب الاشعة.

في هذا العمل، قمنا بتصميم و تطوير نظام ناجح لتصنيف كثافة الثدي الى مجموعتين : دهني و كثيف. واعتمدنا خمس سمات او خصائص لتميز هذه الكثافة : تحليل المكونات الرئيسية (PCA) ونسخة مطورة من هذي السمة تدعى (2DPCA) و SVD و تحليل المصفوفات غير السالبة NMF و سمة العتبه. النتائج كانت مشجعة و تسلط الضوء على طرق جديدة في هذا المجال |

CHAPTER 1

INTRODUCTION

Breast Cancer is one of the most death causes among women worldwide with a percentage of 14% of all cancer types [1]. It occurs in both women and men, but it rarely occurs in men. In addition, Older women have a higher risk of developing breast cancer than younger women [2][3].

Breast cancer is a disease in which abnormal cells in the breast grow faster than the normal ones without control because of Deoxyribonucleic Acid (DNA) damage. It can spread to other parts of the body to produce tumors and replace normal cells. The spread of the cancer, called metastasis, occurs when the cancer gets in the blood or lymph vessels [1].

There are two types of tumors: benign and malignant. Benign tumors are large in size and cannot invade to other tissues. These tumors are not harmful and are not deadly [1]. However, malignant tumors are small in size, dangerous and can cause death [1].

In 2000, the Ministry of Health in Saudi Arabia [4] stated that 2741 new cases of cancer were identified in the Arab countries. In [4], cancer data in Saudi Arabia is contrasted to that in the United States of America (USA). It was found that in the Arab world breast cancer occurs at the age of 52 while it occurs at age 65 in the USA. Moreover, breast cancer is discovered in its late stages in the Arab world.

Zahra Breast Cancer Association [5] reported that in Saudi Arabia about 13.5% of most cancer cases are breast cancer in 2009. The median age at diagnosis was 48 years old. It was also noted that the eastern region in KSA has the highest breast cancer incidence rates.

According to the American Cancer Society [1], the estimation of breast cancer in the USA during 2013 is:

Women:

- 232,340 new invasive breast cancer are identified. Invasive breast cancer can extend from lobules or ducts to cover the surrounding tissue. It can possibly spread into the lymph nodes and other parts of the body. Invasive ductal cancer originates in ducts and invasive lobule cancer originates in lobules.
- 64,640 new cases of in-situ (in place) breast cancer, 85% of these cases are ductal carcinoma in situ (DCIS) and 15% represent lobular carcinoma in situ (LCIS). In-situ cancer is non-invasive breast cancer that can progress to invasive cancer [3].
- 39,620 death cases of breast cancer.

Men:

- 2,240 new cases of breast cancer are identified.
- 410 death cases in men from breast cancer.

Based on previous estimations, death rate decrease of 2% is recorded. This decrease is due to the improvement in breast cancer treatment and early-stage detection.

1.1 Survival Rate

The 5-year survival rate is the normalized number of patients who have cancer and can live at least 5 years. It is based on the number of previous observations of people suffering from cancer. These rates cannot predict the future behaviour of cancer. Instead, these rates are averages that help in knowing the survival chance for patients in similar situations. Table (1-1), shows the 5-year survival rate stages and the percentage regarding each stage [2].

Table 1-1: 5-Year survival rate [3].

Stage	Survival rate
0	100%
I	100%
II	93%
III	72%
IV	22%

If the cancer is detected in the first stages, the patient can be survived or recovered, but this opportunity decreases in the last stages.

1.2 Breast Anatomy

Breast, located in front of the chest, contains mostly fat cells and tissue along with nerves, ligaments, fibrous connective tissue, lymph vessels, lymph nodes, and blood vessels [6]. A female breast is made up of 12-20 lobes. Each lobe contains 20-40 lobules. These lobules contain glands that produce milk. The lobes and lobules are connected to

the nipple through tubes called ducts. Figure (1-1) illustrates the structure of the female breast.

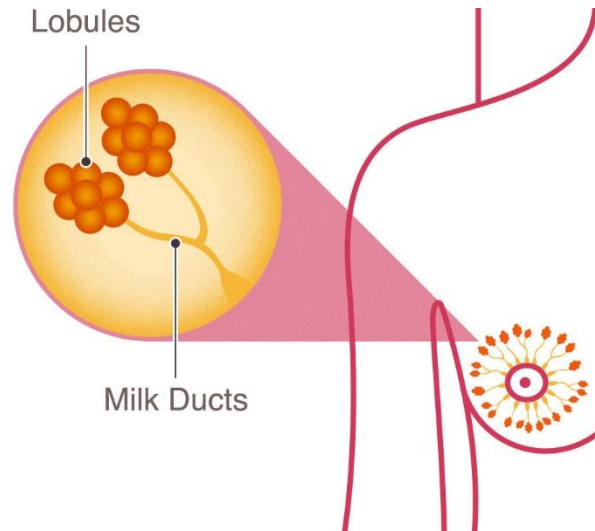


Figure 1-1: Breast structure [6].

Unlike women, men have simpler breast structure. Over time men breasts stop from developing. Furthermore, some milk ducts exist, but still immature and lobules are often absent [6].

1.3 Breast Density

The breast is composed mainly of fat and tissue. In addition, tissue includes the lobules that consist of milk glands, and ducts. In fact, the tissue is the most likely area where breast cancer can start developing [1].

Breast density refers to the amount of tissue and fat in the breast. However, breasts with more tissue than fat are considered dense, whereas breasts with more fat are considered fatty. In general, the tissue appears whiter in the mammography screening. Younger women have denser breasts, which tend to decrease density over time, than

older women. Furthermore, breast density, an inherited characteristic, can be affected by several factors such as age, family history, etc. It can also be affected by changes caused by hormonal fluctuations, including menopause, breastfeeding, pregnancy and menarche.

Several studies [7]–[9] found that dense breasts have a higher risk of developing breast cancer. The breast density can also influence the mammography interpretation [10]. More specifically dense breasts decrease the sensitivity for cancer detection compared to fatty ones. In such cases, cancer regions appear white, and tissue also appears white in mammogram images. In Figure (1-2), a ductal carcinoma in situ (DCIS) appears clearly in a mammography of fatty type according to the BI-RADS category.

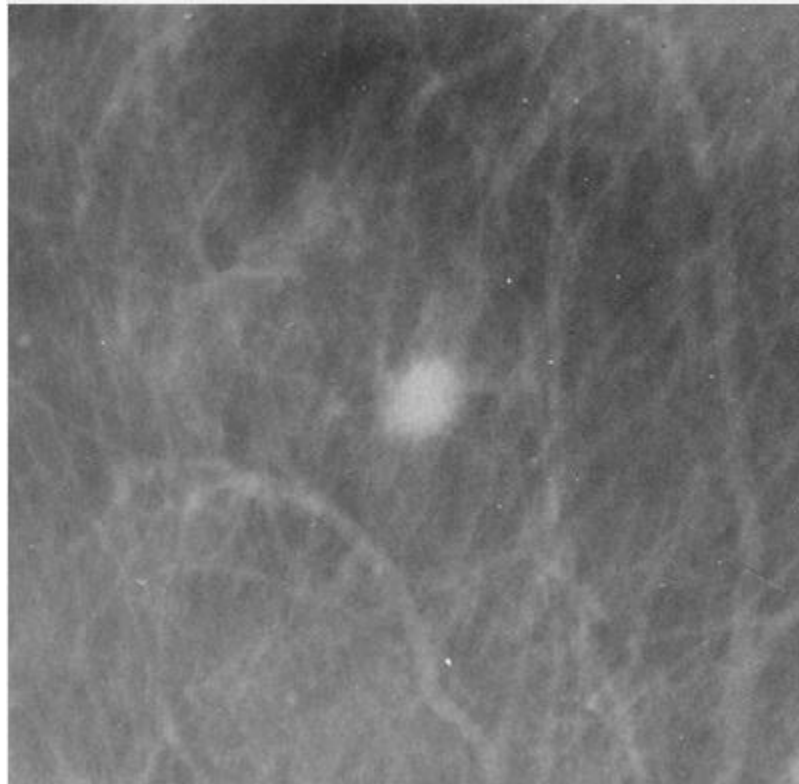


Figure 1-2: A DCIS visible in a fatty dense mammography [11].

1.4 Breast Imaging

Breast imaging aims at early detection of breast cancer. This can help in reducing mortality rates and increasing survival and recovery chances. Also, it helps medical practitioners to decide whether a breast biopsy (operation) is needed or not [3].

There are many types of breast imaging. The common ones are:

1. Mammography imaging.
2. Ultrasound imaging (US).
3. Magnetic resonance imaging (MRI).

1.4.1 Mammography Screening

Mammography screening is one of the methods, recommended by World Health Organization [12], to reduce the mortality rate due to breast cancer. Mammography is one of the best tools used in detecting breast cancer in its early stages.

A mammogram, used by physicians, is an X-ray image that allows checking for breast abnormalities. Also, Detectors are used in digital mammography to convert these x-rays into signals. These electrical signals are used later to produce images that can be processed by the computer [2].

Screening and Diagnostic mammograms represent symptom-free and symptom cases, respectively [2].

Two common abnormalities in the mammogram are masses and calcifications. A mass is the area that occupies the lesion and is shown from two different views. If it can be seen

in a single view only, it is called asymmetric. In addition, Masses occur in different shapes. While benign masses are round and oval with smooth margins, malignant ones come in rough and blur boundary [13].

On the other hand, calcifications are deposits of calcium in the breast. They can be divided into two categories: macro and micro calcification. Micro calcifications, displayed in Figure (1-3), are tiny deposit of calcium. In contrast to micro calcifications, macro calcifications are large deposit of calcium as indicated in Figure (1-4). These macro calcifications are not a sign of breast cancer. However, there is an association between micro calcifications and extra cell activity [13] that relates to tumours. In general, calcifications, gathered in cluster, can be an indicator of a malignant tumour. Moreover, these calcifications are shown in a mammogram as bright dots with different sizes. As a results of this cluster, Benign calcifications are large with smooth boundary, whereas malignant calcifications are small, irregularly shaped with branching on the orientation [14].

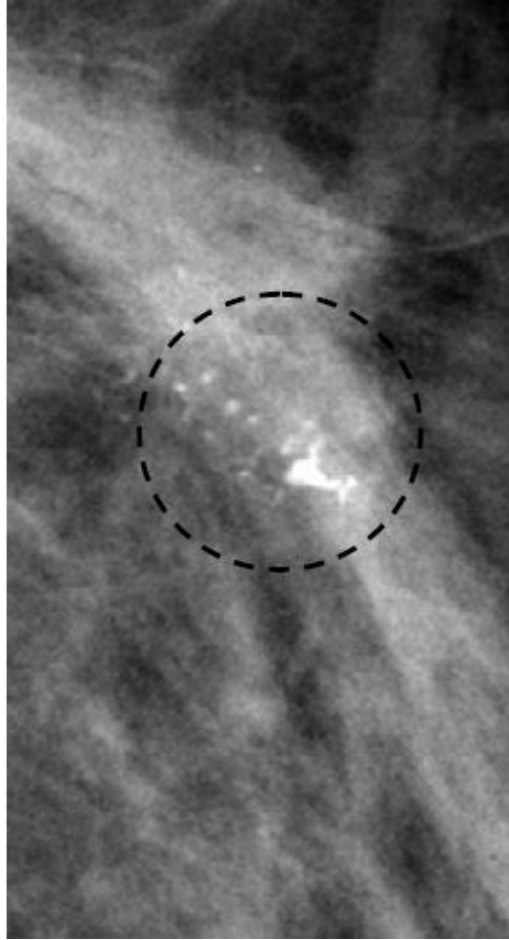


Figure 1-3: Micro-calcification detected on mammogram [13].

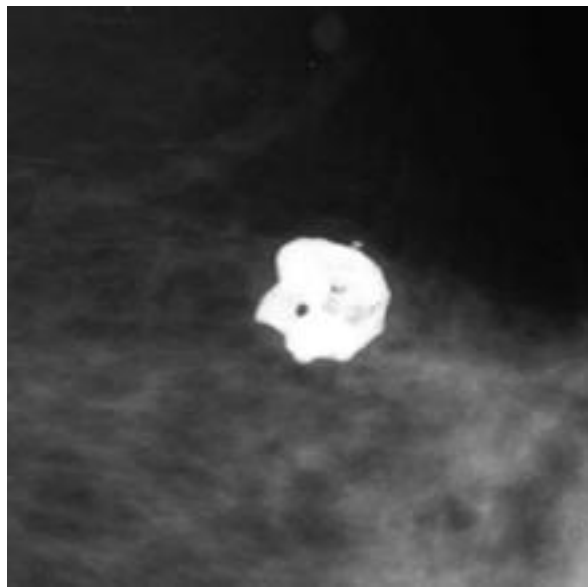


Figure 1-4: Macro-calcification detected on mammogram [15].

It is usual to take different pictures of each breast using different directions and viewpoints to show the inside details. Medio-Lateral Oblique (MLO) and Cranio-Caudal (CC) are the most used viewpoints for mammograms. The CC view is taken from above of the breast and MLO view represent pictures taken from the side of the breast at an angle. Furthermore, MLO view is very important because it allows to depict most of the breast area. In Figure (1-5), the left image is taken using a CC view and the right image represents the MLO view.

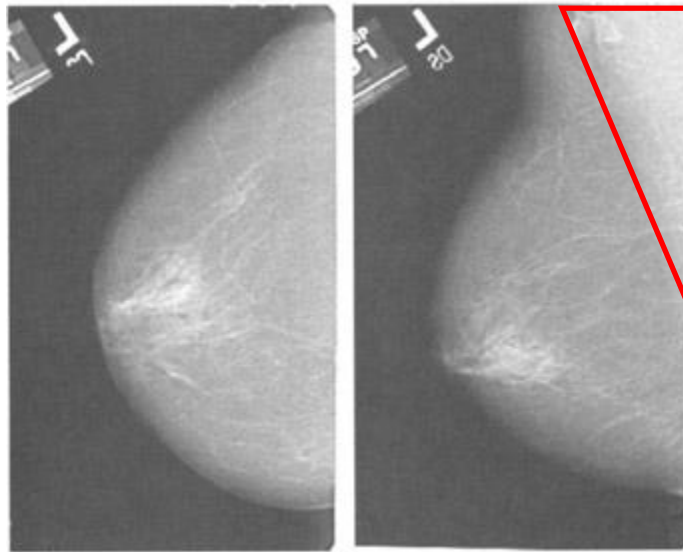


Figure 1-5: CC view (left) MLO view (right) in mammogram screening.

However, the pectoral muscle in MLO views, as clearly indicated in Figure (1-5) where the red triangle illustrates the pectoral muscle, appears in the left or the right upper corner of the image based on the direction in which it is taken. This muscle should be removed in the segmentation phase.

1.4.2 Ultrasound Imaging

Ultrasound imaging can be used to support mammography screening in investigating abnormalities, especially in the case of women with dense breasts. Since it uses sound

waves to create the image, as a result it is safe with no ionizing radiation as in mammography screening. However, some fibrous structures in the breast produce similar acoustic shadowing of breast cancer, which can result in false positive [11]. Usually, ultrasound imaging is not used alone to detect cancer in its early stages since high false positive rates can lead to biopsy and this cause unnecessary harm. In Figure (1-6), an ultrasound image shows a false positive case in which a tissue in the breast is considered as cancer while it is not.

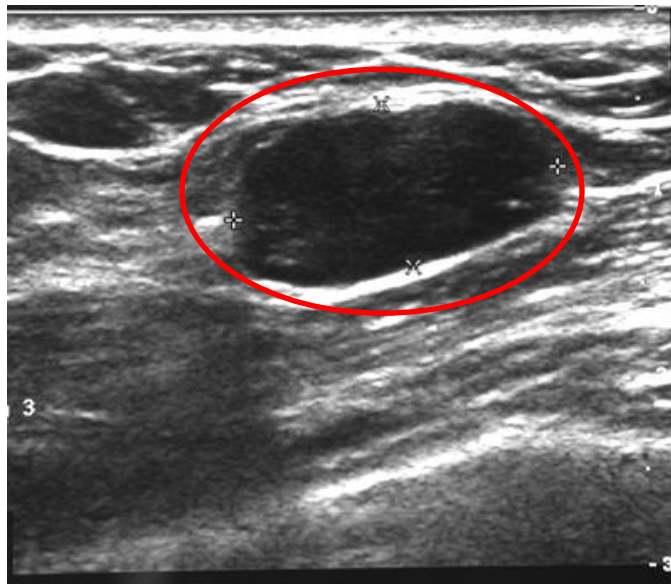


Figure 1-6: A tissue shown as cancer in ultrasound imaging (false-positive) [11].

1.4.3 Magnetic Resonance Imaging (MRI)

MRI is a powerful tool for the evaluation of breast [11]. It provides a good contrast enhancement, which facilitates the easy detection of cancer that is surrounded by fat and also in dense breasts [16]. It is recommended for women with dense breasts* and in case

* Therefore, classifying the breast density is a very crucial step in breast cancer detection and diagnosis

of high risk. Mammogram images and MRI for a dense breast are illustrated in Figure (1-7) where in this case, the MRI image can detect the cancer while mammogram cannot.

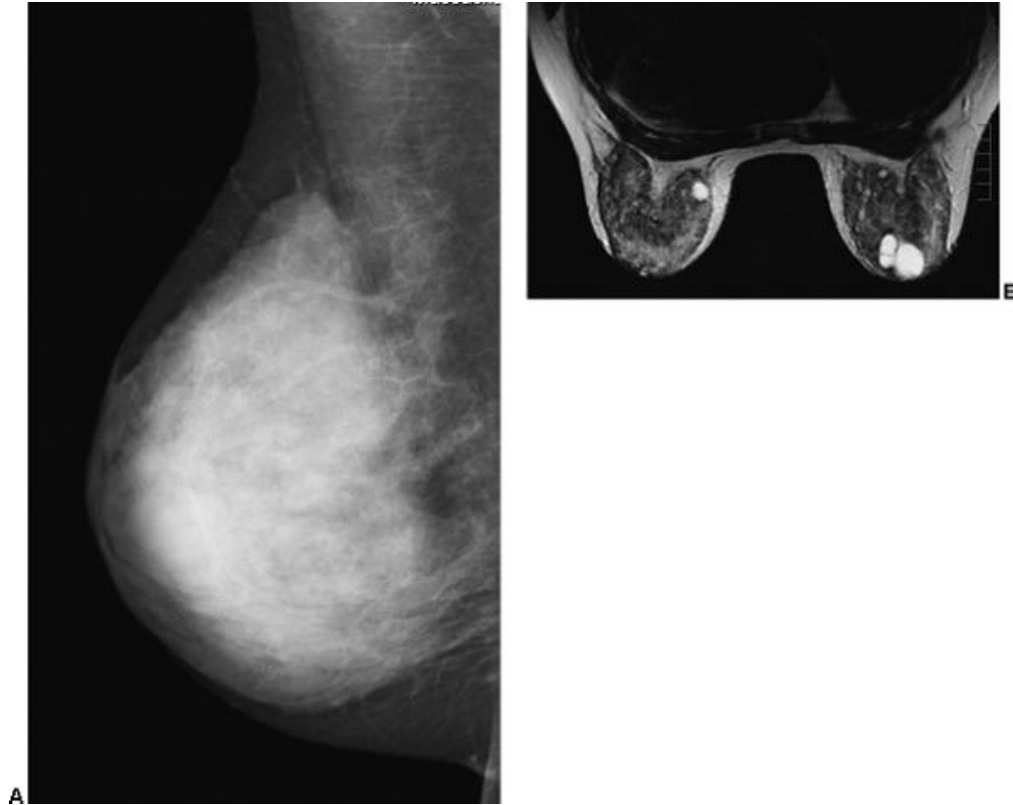


Figure 1-7: (A) Dense breast in mammogram (B) MRI image that shows cysts while mammogram does not [11].

Nevertheless, MRI, like US imaging, produces high false-positive rates. Therefore, it should be used in the advanced stages of diagnosis supporting mammography imaging.

1.5 Breast Imaging Reporting and Database System (BI-RADS)

The BI-RADS nomenclature has been established by the American College of Radiology (ACR) [17] as a standard method for radiologist to describe mammogram reports. The BI-RADS lexicon describes the breast density, lesion feature and lesion classification. Furthermore, the BI-RADS standard classifies the tissue density of the breast into four classes as shown in Table (1-2). An example of the four classes is given in Figure (1-8).

Table 1-2: BI-RADS tissue density classes.

Class	Description
I.	Fatty
II.	Scattered fibro glandular
III.	Heterogeneously dense
IV.	Extremely dense

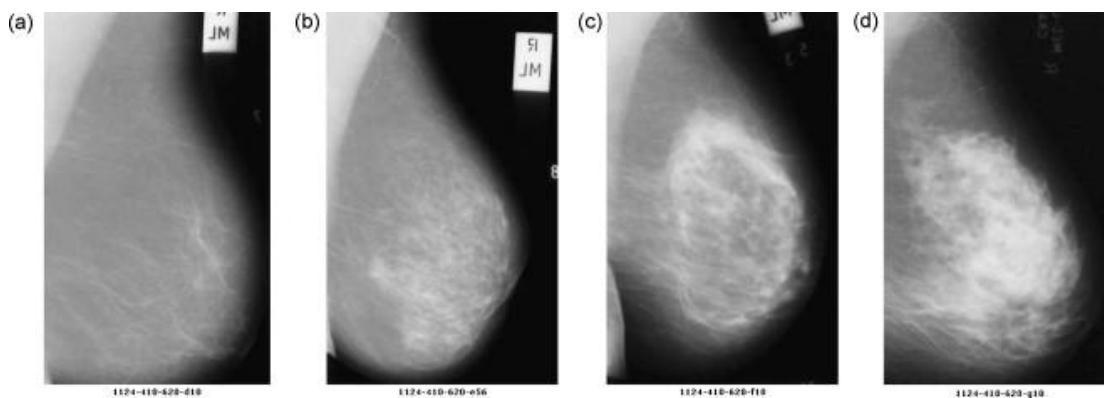


Figure 1-8: Mammograms examples of BI-RADS classes from TYPE I (left) to TYPE IV (right).

The BI-RADS system also defines assessment categories for estimating the lesion and its classification. Table (1-3) shows these categories and the description of each category. This categorization helps doctors and radiologists to record accurate statistics of the patient's case.

Table 1-3: Assessment categories of BI-RADS.

Category	Description
0	Incomplete
1	Negative
2	Benign
3	Probably benign
4	Suspicious abnormal
5	Highly suggestive of malignancy
6	Proven malignancy

1.6 Computer-Aided Diagnosis (CAD)

Radiologist evaluates mammograms based on their visual analysis. The misinterpretation can lead to more false-positive cases and biopsies, which turn out to reveal benign tumours. More specifically, about 65-90% of biopsies turn out to be benign [18]. Computer-aided Diagnosis (CAD) systems appear to be helpful for the radiologist in their examinations in detecting breast cancer in mammograms and assist in choosing between follow-up test and biopsy [13]. Also CAD systems decrease the variability in readings of radiologists, and therefore leads to more precise diagnosis decision and decrease the number of false-positive rates [13].

CAD systems involve many image processing algorithms. These algorithms consist of standard steps presented in Figure (1-9). Pre-processing, the first step, removes the noise from digital mammogram and improves the quality of the digital image. In mammograms, background and pectoral muscle must be removed in the case of MLO views. The segmentation step finds the suspicious region of interest (ROI) that contains the abnormalities [13]. Features are calculated in the feature extraction step based on the characteristics of ROI. In the feature selection step, a number of the extracted features are

selected which provide high classification accuracy and reduce false positive rate. Finally, breast cancer or density classification is performed in the classification step.

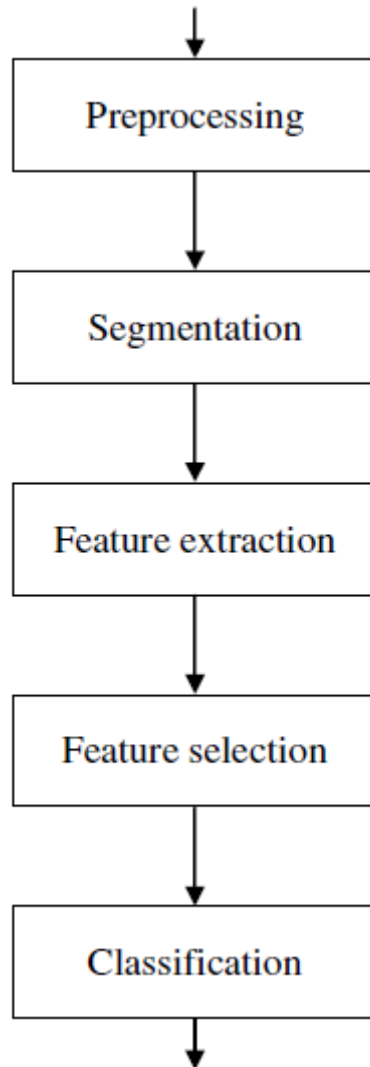


Figure 1-9: Image processing typical steps in CAD systems.

1.7 Classification Performance Measures

In pattern recognition and machine learning applications, the confusion matrix is used to measure the performance of the classification algorithm. The confusion matrix is a table where columns represent the predicted class and rows represent the actual class. In the

case of binary classification problems the confusion matrix looks as indicated in Table (1-4).

Table 1-4: Confusion matrix in binary classification.

		Predicted Class	
		Yes	No
Actual Class	Yes	TP	FN
	No	FP	TN

The possible outcomes in the binary classification case are ‘true positive’ (TP), ‘false positive’ (FP), ‘true negative’ (TN) and ‘false negative’ (FN). A false-positive occurs when the sample is classified incorrectly as positive while it is negative. Classifying a sample as negative when in fact it is positive causes false-negative. True-positive and true-negative are correct classifications of positive and negative samples, respectively.

Using the confusion matrix, the accuracy, precision, true-positive rate (TPR) and false positive rate (FPR) can be calculated using:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1.1)$$

$$precision = \frac{TP}{FP + TP} \quad (1.2)$$

$$TPR = \frac{TP}{FN + TP} \quad (1.3)$$

$$FPR = \frac{FP}{TN + FP} \quad (1.4)$$

Using similar concepts, the area under the curve (AUC) of the receiver operating characteristics (ROC) indicates the performance of the classifier. Given the normalized measures, TPR and FPR measures, the total AUC are equal to one. Larger AUC values correspond to better the classifier. In addition, the ROC curve is defined by plotting the TPR measure versus FPR.

1.8 Problem Statement

Medical imaging, as an analytical method, is a key tool for the inspection of the internals of the human body. Several modalities allow radiologists to examine the internal structure and these modalities receive a great interest in several researches. Each of these modalities has a great importance in certain medical domain.

Mammography is one of the best methods used in early breast cancer detection. Obviously, Breast Density is one of the best indicators of breast cancer. Although CAD systems automate the process of breast density classification, these systems still need more improvements. From this motivation, the aim of Breast density classification system is to help radiologists for evaluating mammography for breast cancer detection.

The main objective of this thesis is to use different machine learning techniques and introduce new techniques used in classifying breast density in digital mammograms according to BI-RADS lexicon. These techniques will be discussed in later chapters.

1.9 Thesis outline

This thesis is organized as follows: Chapter 2 provides a detailed account of the literature review of the methods used in breast density classification. In Chapter 3, the proposed

classification approach is described in details. Furthermore, the database and tools used in this work are discussed in Chapter 4. The system is described in details, implements different features and uses SVM and KNN as classifiers. Moreover, the performance results of the developed system are discovered in this chapter. Finally, in Chapter 5 the conclusions are given and future work directions are outlined.

CHAPTER 2

LITERATURE REVIEW

In this chapter, relevant state-of-the-art techniques and methods are reviewed and summarized. Some of these techniques are deeply related to the proposed approach. The techniques can be divided into Quantitative and Qualitative approaches based on density classification.

2.1 Quantitative Approaches

Quantitative approaches approximate breast density and express it as a percentage. Many approaches are proposed in this category. Yaffe provides a detailed survey on these methods [19]. Cumulus and Interactive threshold methods [20] are commonly used in clinical studies. This approach, based on a threshold defined by the user, is applied on digital mammograms. The user selects different thresholds to identify several areas in the image. To determine the amount of density, the histogram is computed for the whole segmented breast area and the dense area. Thus, the ratio between these quantities represents the density. This semi-automatic approach is user dependent which can produce some variability from one user to another.

Also, volumetric assessment methods are widely used. These methods find the volumetric density from a 2D digital mammogram. Highnam et al. [21] provide an explanation to a new approach, VolparaTM that finds a fat area. Based on this area, thickness of each pixel in the mammogram image is calculated. The integration of pixel

thickness values of these pixels represents the volumetric density of the image. To verify the validation of this method a comparison is made between Volpara and cumulus [22].

This comparison shows that they are all closely related.

A similar method to cumulus, that identify dense tissue in the mammogram, is proposed by [23]. Unlike cumulus, this method can recognize the regions in the breast automatically and can find an optimal threshold between fatty and dense tissue. This method shows similar results compared to cumulus.

Chen et al. [24] proposed a quantitative measure that use a topographic map to represent the breast tissue density. In this approach, a connected component represents a shape is constructed as tree that describe the topological structure. To detect dense regions, the saliency and independency features are used. MIAS and DDSM databases are used in the evaluation and the obtained accuracies are 76% and 81% respectively.

2.2 Qualitative Approaches

Instead of representing the density as a percentage value, these approaches divide the density into several categories, such as BI-RADS categories. The approaches in this literature can be grouped into three categories:

- 1- Matrix Factorization Methods
- 2- Global Histogram Methods
- 3- Texture Analysis Methods

2.2.1 Matrix Factorization Methods

The matrix factorization techniques decompose a data matrix into a product of several matrices according to different constraints. Moreover, these techniques help in reducing the dimension of the data. In the breast density classification the data matrix that contains the mammogram images have a very high dimension and using these techniques can help in reducing the dimension of the data matrix. However, there are several factorization algorithms, each utilizes certain constraint that results in different representation properties.

Oliver et al. [25] have proposed system for the segmentation of mammogram and for classifying the breast density into fatty and dense. For breast density classification they use Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) as features. They obtain better results with PCA.

PCA works in 1D vector, so the image is converted into 1D vector. Unlike PCA, 2D PCA is an extension of PCA that deals with 2D matrices. It cuts the computation cost of the standard PCA.

Consider an image A as $m \times n$, and x as N dimensional unit column vector. Projecting A into x results in an M dimensional vector y where $y = Ax$.

2DPCA finds a good projection vector x by tracing the covariance matrix of the projected feature vectors. The covariance matrix can be obtained by adopting $J(X) = \text{tr}(S_x)$.

The covariance matrix of A is given as

$$S_x = E[(A - EA)^T(A - EA)] \quad (2.1)$$

De Olivera et al. [26] uses 2DPCA as feature to classify fatty and dense mammograms in a Content-Based Image Retrieval (CBIR) system. Taking the first 4 principle components as features with Support Vector Machine, with Gaussian kernel, as classifier they show that 2DPCA is more accurate than standard PCA. Deserno et al. [27] extend the work by increasing the number of classes to cover the 4 BI-RADS classes with the same feature and classifier.

Other approaches use Singular Value Decomposition (SVD) technique. De Oliveira et al. [28], proposed a CBIR model called MammoSVD. The system is developed to classify the density of the breast – fatty or dense using SVM. The obtained SVD values provide useful information of image texture. SVD is also used in the reduction of dimensionality. The goal is to find a sufficient rank k of singular values that can improve the characterization of the image. This value must be minimum than the dimension of the data. In addition, the factored matrix represents the intensity of the pixels that belong to a certain texture in the image. Using SVM the system achieves an accuracy of 90%. In MammoSVx [29], another CBIR system, the 4 BI-RADS classes are considered and the system is able to achieve 82.14% with 25 singular values by taking and passing these features to SVM classifier with polynomial kernel. The database used, contains 10000 mammogram images taken from different sources.

2.2.2 Global Histogram Methods

Extracting features from the global histogram is addressed by Sheshadri [30]. From the MIAS database, six statistical features are extracted. These features are shown in Table (2-1).

Table 2-1: MIAS-based statistical features [30].

Feature	Expression
Expectation	$m = \sum_{i=0}^{L-1} z_i p(z_i)$
Standard deviation	$\sigma = \sqrt{\sigma^2}$
Smoothness	$R = 1 - \frac{1}{1 + \sigma^2}$
Skewness	$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$
Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$
Entropy	$e = - \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$

This approach obtains 80% accuracy validated by expert radiologist. Instead of six histogram moments Liu, Li et al. [31] use three higher order histogram moments. A preprocessing phase is applied to exclude noise. After excluding noise, a dyadic wavelet decomposition is performed. The three resolution levels of approximation images were calculated. The higher order central moments up to the fourth order determine the histogram variance, skewness and kurtosis. From these components the density feature vector \mathbf{v} is extracted. From the density feature vector 18 features are calculated (six features for each component) as follows:

$$v_{1j} = \sum_{i=0}^{L-1} (z_i - m_j)^2 p_j(z_i) \quad (2.2)$$

$$v_{2j} = \sum_{i=0}^{L-1} \frac{(z_i - m_j)^3 p_j(z_i)}{v_{1j}^{3/2}} \quad (2.3)$$

$$v_{3j} = \sum_{i=0}^{L-1} \frac{(z_i - m_j)^4 p_j(z_i)}{v_{1j}^2} \quad (2.4)$$

where $\mathbf{j}=0, 1, 2$ represents the \mathbf{j} th resolution, $\mathbf{p}(\mathbf{z}_i)$ is the probability of the \mathbf{i} th bin and the gray value of the \mathbf{i} th bin determined by \mathbf{z}_i , L determines bins count in the histograms, \mathbf{m}_j is the average of the intensity.

For y direction, the histogram considered as:

$$v_{4j} = \frac{1}{L} \sum_{i=0}^{L-1} (p_j(z_i) - p_{mj})^2 \quad (2.5)$$

$$v_{5j} = \frac{1}{L} \sum_{i=0}^{L-1} \frac{(p_j(z_i) - p_{mj})^3}{v_{4j}^{3/2}} \quad (2.6)$$

$$v_{6j} = \frac{1}{L} \sum_{i=0}^{L-1} \frac{(p_j(z_i) - p_{mj})^4}{v_{4j}^2} \quad (2.7)$$

DAG-SVM [32] is used to classify the three groups. This approach uses small dimension which reduces the overhead computations. It shows better results in the two classes (fatty or dense) classification, but it needs an improvement in the case of multiple classes.

Oliver et al. [33] discuss a segmentation and classification system. For breast density classification, features are extracted from the co-occurrence matrices, which is a two-dimensional matrix of histograms. These histograms are the occurrence of gray-levels pairs for a displacement vector. A matrix $P_{ij}(d, \theta)$ of relative frequencies that specify the co-occurrence of gray levels, in which a distance d between two pixels and angle θ contains the gray levels i and j [33]. This method uses 4 different direction angles, which are 0° , 45° , 90° , and 135° ; with a distance of 1. The contrast, entropy, energy, average, correlation, difference average, entropy, homogeneity and difference entropy features are determined for each co-occurrence. For classification purposes, this approach has two classifiers: K- Nearest Neighbour and a Decision Tree classifier. These two classifiers achieve better results when combined together. The average is taken in the final result. The results show an accuracy of 47% when combining the two classifiers, 43% for the ID3 that determine the threshold of the decision tree and 40.3% for k-NN with the use of 300 right side MLO mammograms taken from DDSM.

An automatic breast density classification system is described by Oliver et al. [34]. This system segments and classify mammogram densities into fatty and dense. Extracting four histogram moments as morphological features and a set of texture features from the co-occurrence matrices. The system is able to obtain up to 83% using Bayesian combination of different classifiers, in this case Decision Tree and KNN Classifiers.

Matsubara et al. [35] provided an approach that segment the mammogram into three regions using variance histogram analysis and discriminant analysis. The Fibro glandular tissue density is classified based on a ratio obtained from these regions. They use 148

Japanese digitized mammograms evaluated by expert radiologist. This system obtains 90% accuracy based on four categories.

Subashini et al. [36] use statistical features and histogram moments as features in mammogram density classification. Providing these features to SVM classifier, this approach achieves 95% accuracy on the MIAS database.

Another system was proposed by Sheshadri et al. [30]. They use statistical features as features with the aid of expert radiologist in the evaluation, the system achieved 80% accuracy on MIAS database.

Petroudi et al. [37] presents an approach that use a set of Amplitude-Modulation Frequency-Modulation (AM-FM) features. The instantaneous amplitude (IA) component is extracted using different filters of scaling and band passing. From the maximum IA, the normalized histogram used to represent the breast densities. In classification the k-nearest neighbour was used with $k=5$. This approach is evaluated on MIAS database and the results reaches 84% accuracy.

A hybrid classification method proposed by Sharma et al. [38], that uses correlation-based feature selection (CFS) and sequential minimal optimization (SMO) to classify fatty and dense mammograms. The dense in mammogram is modelled using several texture features. These features are reduced using CFS. Using the MIAS database, the SMO classifier correctly classified 96% of mammograms.

Vallez et al. [39], proposed an approach that uses histograms, texture information of the 2nd-order statistics of histograms, and space frequency properties as features for the four BI-RADS categories. Dimension of features reduced using PCA and LDA. A voting tree classification classifier is proposed that combine several classifiers such as Support

Vector Machine, Neural Networks, and K-Nearest Neighbour. The reported results of the experiment shows that 91% correctly classified on a database of 1137 images, while 99% recognition rate obtained using MIAS database (322 samples).

2.2.3 Texture Analysis Methods

Texture analysis techniques have a variety of application domains. They can also be used in breast density classification [40]–[42]. The texture is a set of patterns that may or may not have a well-defined structure [43]. It can describe several surface characteristics. The image texture is a collection of primitive units in a regular or iterated pattern [44]. Statistical approaches can help in analyse the structure of the texture. It can identify the texture and represent it with a quantitative measure of intensities in a region [44]. This texture information is associated with the variation in the intensity of the image and it can be seen as a function of surface texture, orientation, illumination, view of the camera and hence any change in these attributes may result in a variation in texture [44] [45]. Statistical methods examine the spatial distribution of gray level values. This is done by calculating local features for each pixel of the image and extracting a collection of statistics from the distribution of the local features [46].

Textons based dictionary are one of the widely used techniques in analysing textures [40]. They are a fundamental block for texture, which can be used to extract the characteristics of the mammograms. Textons are defined as the representative responses occurring after convolving an image with a set of filters, “filter banks” [44]. Leung and Malik also defined Textons as clustered filter responses [47]. Most texton dictionary based approaches have two stages of training. In the learning stage, the filter responses

are used to describe the local structure in the images [48]. A filter is an $N \times N$ matrix convoluted with the pixels of the image results in different features of pixels [44] [45]. The most used sizes are 3×3 , 5×5 , 7×7 , 25×25 and 49×49 and the most used filter in classifying breast density is Maximum Response 8 (MR8) filter [44]. This filter bank consists of 38 filters at various orientations and scales. The filters are shown in Figure (2-1). There are 36 first and second derivative of Gaussian at six orientations and three scales, and Gaussian and Laplacian of Gaussian used directly. The rotation invariance is achieved by measuring the maximum response across orientations only. Thus, maximum response reduces the number of responses from 38 to 8. The MR8 filter bank contains 38 filters, but 8 responses only.

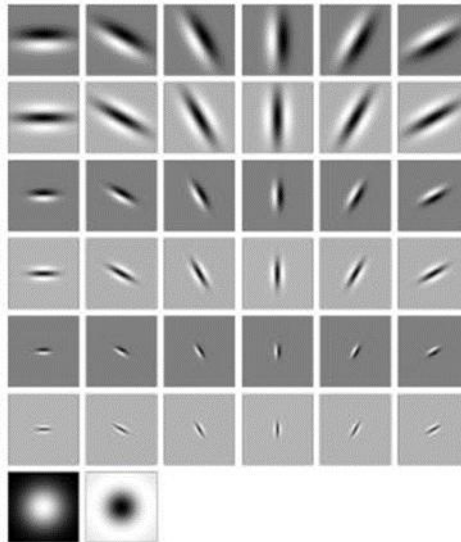


Figure 2-1: MR8 filter bank.

After applying the filters, the texton dictionary is constructed from the filter responses that are grouped using K-means clustering algorithm. A K centers for each class are

chosen and referred to as textons. So the number of textons depends on the number of centers. Figure (2-2) represents the learning stage.

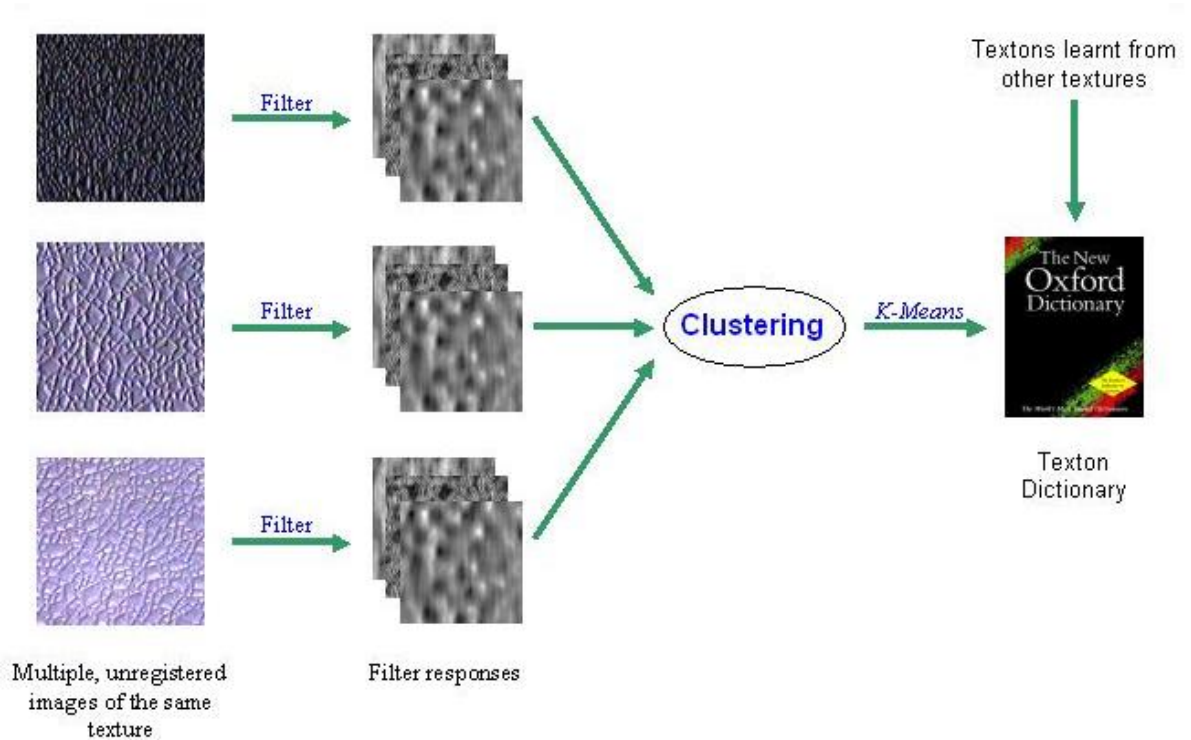


Figure 2-2: Textons Learning stage [49]

The next stage after learning the dictionary is to learn models for each class. Filter banks are applied to the training images and each filter response given a label by closest texton. The texton histograms are computed and a set of histograms represent a models for the breast density classes. Figure (2-3) shows the modelling step.

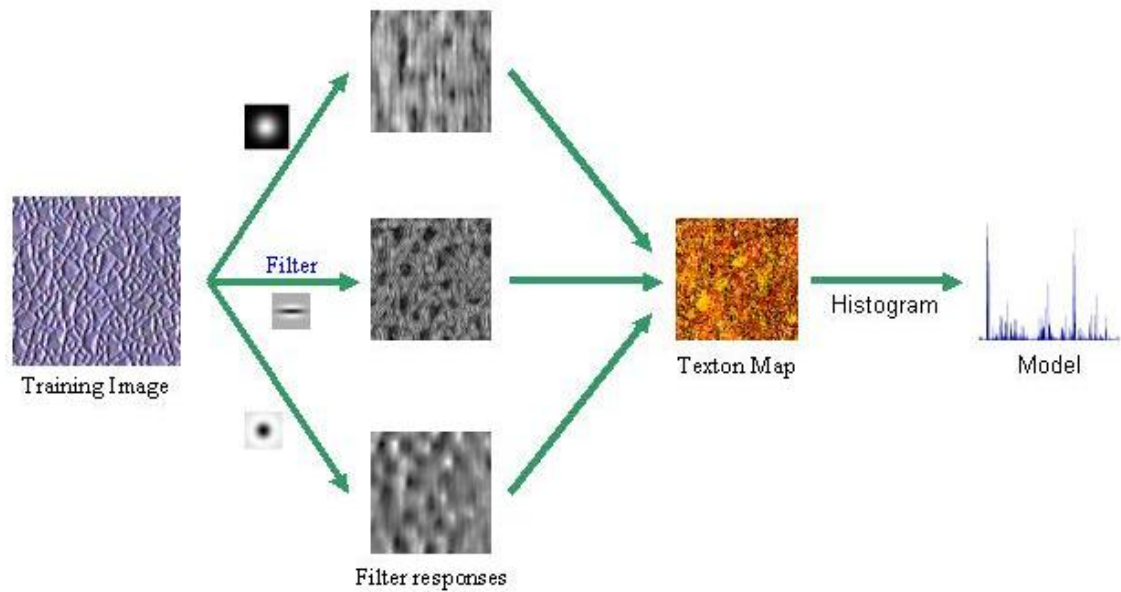


Figure 2-3: Texton Modeling Stage [49]

In the classification stage, the same steps are applied to the test image. After obtaining the histogram of the test image a comparison with the learnt models are performed.

Petroudi et al. [42] use texton spatial dependence matrix in regions of breast that correspond to texton map. In this approach a new structural and statistical texture information is introduced called texton co-occurrence matrix or Texton Spatial Dependence Matrix (TSDM). This matrix contains the frequencies of the textons co-occurrences. Using Oxford mammogram database they obtain 82% for the four BI-RADS classes using chi-square distance measure and 90% for binary classification considering two BI-RADS classes fatty and dense.

Other interesting local texture features used are Local Binary Pattern (LBP) and Scale Invariant Feature Transform (SIFT).

LBP of texture intensities provides a robust way for describing clean local binary patterns insensitive to changes in illumination with low computational complexity.

The SIFT technique transforms an image into a collection of local features. These features are invariant to rotation, translation and scaling. The algorithm first convolve the image with Gaussians. From the smoothed images a differences of Gaussians are generated. Scale space extrema detection stage finds points of interest as local extrema. The gradient histogram is computed from key point and key point descriptors. From the key point orientation the feature vector is extracted which contains orientation histograms on 4x4 pixel neighbourhoods.

In [40], Chen et al. use LBP, texton, Local gray level appearances (LGA) and Basic Image Feature (BIF) where MIAS database is used in the experiment. Using KNN as classifier they obtain maximum of 75% using Texton for the four BI-RADS, for the two BI-RADS classes (fatty and dense) they reported up to 88%.

Liasis et al. [41] extract SIFT, LBP and texton features from MIAS database, combining these features together they reported recognition rate 93.4% for 3 classes using SVM classifier.

Bosch et al. [50] propose a methodology to classify breast parenchymal tissue density. Their approach has two steps, first the tissue density distribution is discovered through unsupervised algorithms. In this phase, they studied SIFT and textons as features. The second step involves probabilistic Latent Semantic Analysis to classify breast densities

according to BI-RADS system. Results show that texton outperform SIFT. The reported accuracy of the system achieved 91% using MIAS database.

Tzikopoulos et al. [51] presented a methodology to segment and classify mammograms density. After segmenting the mammogram and applying pectoral muscle removing algorithm, a breast density classification procedure is applied. In this procedure they use a new fractal dimension as a feature and support vector machine as a classifier. Using this approach they achieve an accuracy up to 85.7 via MIAS database.

Liu et al. [52] provided a methodology for using wavelet transform to find sub-regions. From the sub-regions the histograms are used to model densities. These histogram features are passed to SVM classifier. The reported accuracy of the system was 86%.

2.2.4 Summary of the literature

The techniques in the literature are summarized in Table 2-2.

Table 2-2: Summary of literature review.

Year, Ref	Database	No. Images	ROI Size	Features	Classifier	Reported Accuracy	No. Classes	Comments
2001 [35]		147		Histograms	Expert	90%	4	Local Database
2005 [33]	DDSM	300		Histogram form Co-occurrence Matrix	KNN + Decision Tree	47%	4	ROI not specified
2006 [50]	MIAS	322	1024x1024	SIFT + Texton	pLSA	91%	3	

2007 [30]	MIAS	322	1024x1024	Histogram Moments	Expert	80%	2	
2009 [28]	IRMA	800	1024x500	SVD	SVM	90%	2	
2010 [25]	MIAS + Trueta			PCA + LDA		> 89%	2	No. and size of images not specified
2010 [26]	IRMA	1392		2DPCA	SVM	80%	2	Size not specified
2010 [31]	MIAS	322	1024x1024	Histogram moments	DAG- SVM	80-77%	2-3	Accuracy for 2 and 3 classes respectively
2010 [36]	MIAS	322	1024x1024	Histogram moments	SVM	95%	3	
2011 [27]	IRMA	9870	128x128	2DPCA	SVM	80%	4	
2011 [52]				Wavelet transform	SVM	86%	4	
2011 [29]	MIAS + DDSM + LLNL + RWTH	10000		SVD	SVM	82%	4	
2011 [40]	MIAS	322		LBP + Textons + LGA + BIF	KNN	86%- 75%	2-4	Accuracy for 2 and 4 classes respectively
2011 [42]	Oxford	100		TSDM	chi- square distance	90%- 82%	2-4	Accuracy for 2 and 4 classes respectively
2012 [41]	MIAS	322	1024x1024	SIFT + LBP + Textons	SVM	93.4%	3	

2013 [37]	MIAS	322	1024x1024	Global histogram	KNN	84%	3	
2014 [38]	MIAS	322	1024x1024	Texture features	SMO	96%	2	
2014 [39]	MIAS+FFDM	1459	1024x1024	Histograms	Voting Tree	99-91%	3-4	Sizes of FFDM haven't specified

CHAPTER 3

PROPOSED SOLUTION

It is apparent in the literature that a lot of research has been undertaken in the mammogram density classification. However, there is a need for more revision to improve the accuracy of these systems. The main aim of this thesis is to classify breast density according to BI-RADS lexicon using machine learning techniques. This chapter provides an explanation about the features and classifiers used. It also describes the proposed solution, discusses each phase in details.

3.1 Methodology

To achieve the objectives of this thesis in classifying breast density based on BI-RADS, a density classification system is developed. The major components of the proposed system are shown in Figure (3-1).

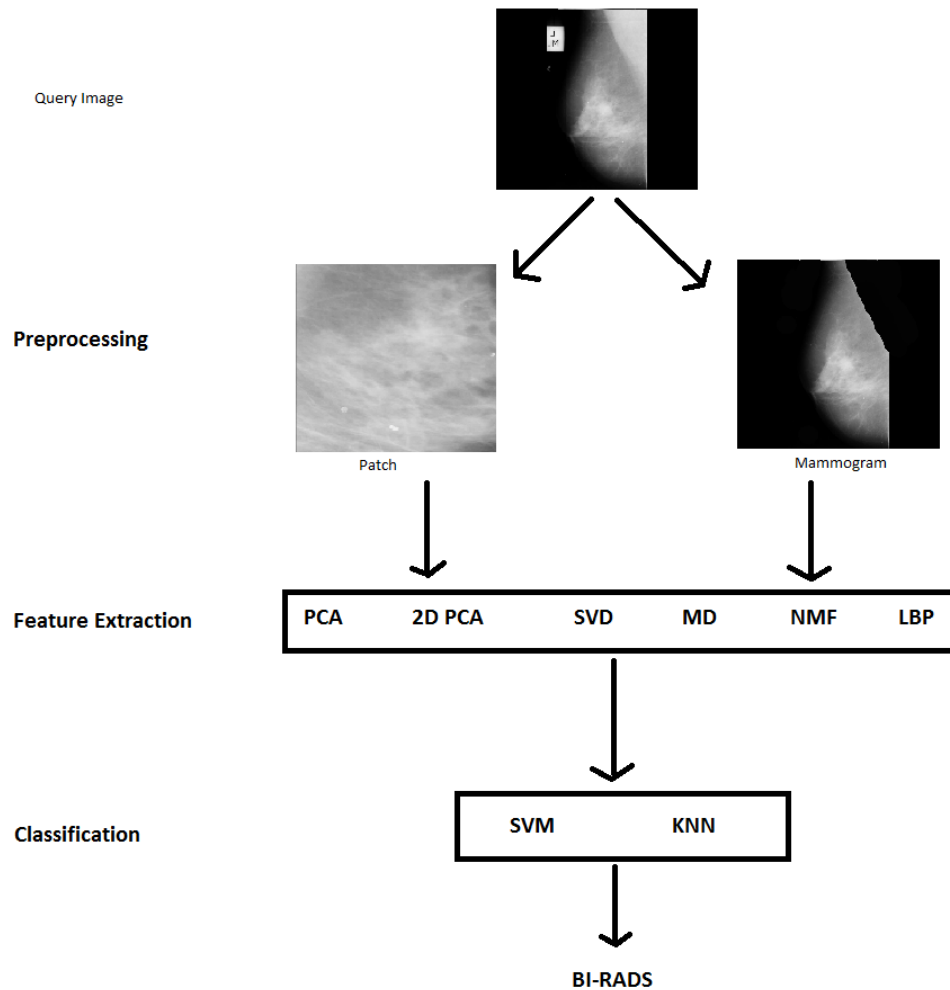


Figure 3-1: Proposed breast density classification system.

First a query image needs a pre-processing step to remove labels and pectoral muscle. Moreover, ROI of size 300x300 is also used. An explanation of this step will be discussed later in section 3.1.1. In addition, Mammogram and ROI extracted to check which one improve the system performance. After pre-processing, the features: PCA, 2D PCA, SVD, Mean Density, LBP and NMF, are extracted and passed to the classification step which include SVM and KNN classifiers, since these classifiers are commonly used in the literature to determine the category of the mammogram density.

An explanation about these steps is provided below:

3.1.1 Preprocessing and Segmentation Step

Interpretation of Mammography is not easy as thought, so the pre-processing step is very important in reducing the noise in mammogram, such as labels to make the feature extraction more reliable. The segmentation excludes the pectoral muscle in case of MLO view and extract ROI from the mammogram. Since mammograms segmentation is out the scope of this work, this step is done manually by [53] and all the images are available in the web. Using MIAS database, the labels are cleaned and the pectoral muscle is extracted from the mammograms. In addition, they provide ROI of size 300x300 pixels for all the images in the MIAS database.

In Figure (3-2) the image 'mdb004.pgm' sample shown to the left where the labels and pectoral muscle appears clearly in the image and the segmented image shown to the right.

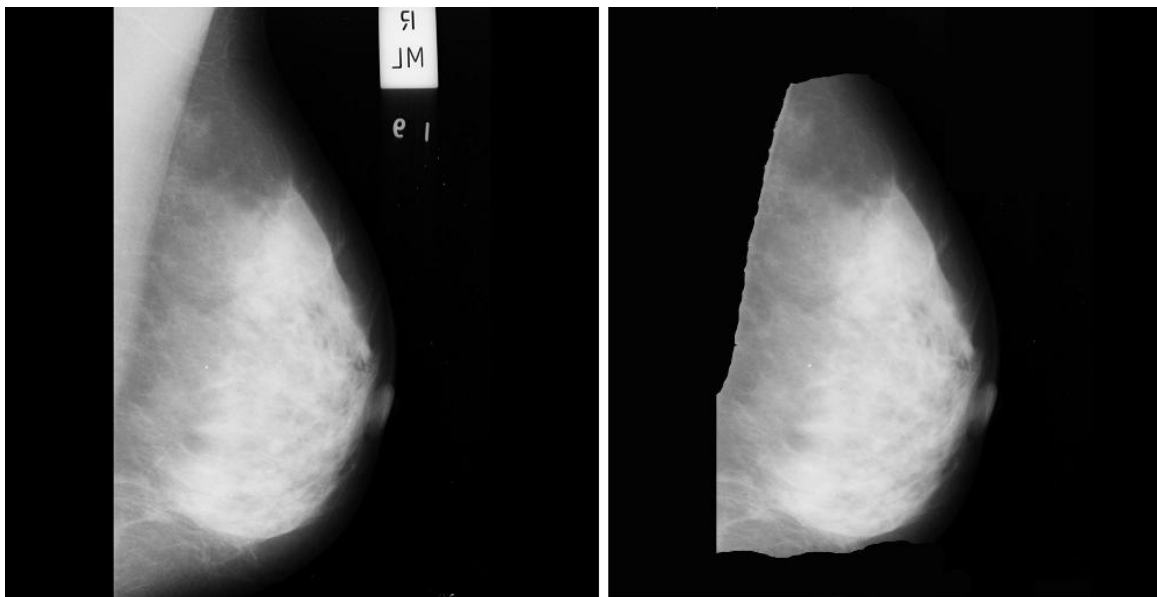


Figure 3-2: Extracting pectoral muscle and labels from a mammogram. (left) original mammogram (right) segmented mammogram.

An ROI (patch) taken from the same sample above, after removing noise, is presented in Figure (3-3).

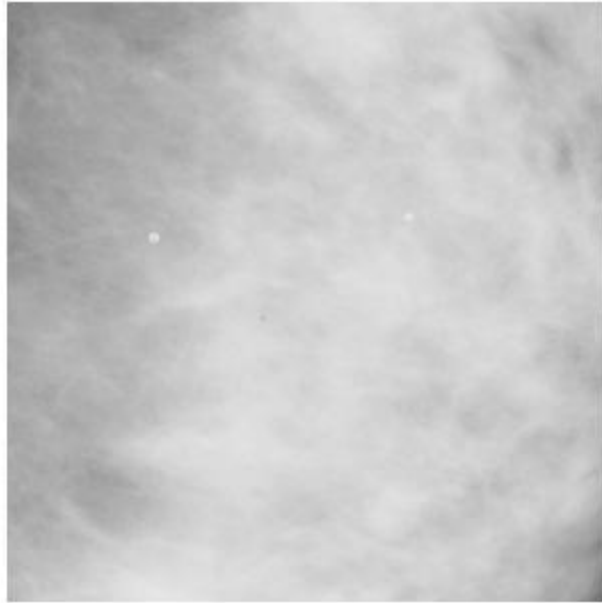


Figure 3-3: A 300x300 pixels mammographic patch.

3.1.1.1 Database Downsizing

Some of the used features here require more memory in their calculations. When PCA and SVD features are extracted from database images of size 1024 x 1024, they produce a huge covariance matrix (the size is more than 1,000,000 x 1,000,000) which require a huge computation power. So to reduce this computations while retaining the image details, the database images are downsized. This can be done while considering the Peak Signal to Noise Ratio (PSNR) similarity measure between the actual image and the resized image.

PSNR similarity measure is commonly used for quality measure of lossy image compression codec. It uses logarithmic decibel scale (dB) to express the perceptual

distortion between two images. However, this distortion can be caused by downsizing process. In PSNR, a higher value indicate good similarity or quality.

To find the PSNR value, first the Mean Square Error (MSE) is calculated using the following equation:

$$MSE = \frac{\sum_{M,N}(I_1 - I_2)^2}{M * N} \quad (3.1)$$

M and N are the number of rows and columns of the images I_1 and I_2 . Then the PSNR is calculated as:

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE} \right) \quad (3.2)$$

Where R is the maximum possible pixel value. For example, if the image is represented as 8-bit then R=255. However, the typical value of PSNR for 8-bit images for human vision perceive is 30 dB.

The following procedure is used when downsizing the database images:

1. Read the image X from the database.
2. Resize the image to the desired size and store in Y.
3. Reconstruct the original image from the resized image Y.
4. Calculate the PSNR between X and Y.
5. Repeat these steps until all images in database are read.
6. Find the average of PSNR values.

Table (3-1) shows the PSNR values (for all MIAS images of 1024x1024 pixels) for different sizes.

Table 3-1: PSNR values for different sizes of MIAS mammograms.

size	dense	fatty-glandular	fatty
100x100	34.54	34.32	34.08
200x200	38.18	37.75	37.12
300x300	40.49	40.12	39.63
400x400	42.67	42.23	41.71
500x500	44.06	43.7	43.13

Calculating PSNR values for mammogram patches, the results are shown in Table (3-2).

Table 3-2: PSNR values for different sizes of MIAS patches.

size	dense	fatty-glandular	fatty
100x100	44.27	42.67	42.5
200x200	48.8	47.44	47.24

From Table (3-1) and (3-2), the size 100x100 is selected for downsizing the database as it has an acceptable PSNR value.

3.1.2 Feature Extraction

As shown in the literature, there are a number of features that are used to describe the texture of the mammograms. In order to classify breast density, a detailed breakdown of the used features is given.

3.1.2.1 Principle Component Analysis (PCA)

PCA is a successful technique from a family of techniques that take highly dimensional data, and use dependencies between variables to represent it in a lower dimension without losing too much information. It is also called Karhunen-Loeve transformation (KLT).

The PCA technique is based on finding a desirable number of principle components (the directions of the new sub-space) of multidimensional data. These principle components can be derived by many ways. However, the simplest method is to find a projection that maximizes the variance. So the first principle component is the one with largest variance and the second is the one that has the second largest variance and is orthogonal to the first principle component, and repeat until all principle components are calculated.

The key point in PCA is to calculate the eigenvalues and eigenvectors of the covariance matrix since the covariance matrix tells us information about relationship of data elements, if they increase, decrease, or independent. Using such factorization, enable us to extract lines that characterize the scatter of the data.

Eigenvalues or characteristic roots are scalar values associated with matrix equation. Each eigenvalue is paired with an eigenvector.

In Linear Algebra, Diagonalization of a matrix is the process that takes a square matrix and converts it into a diagonal matrix that has the same characteristics of that matrix. Finding the diagonal matrix is the same as finding the eigenvalues.

$$A = USU^{-1}$$

Here, the matrix A has been factored into three matrices: U is matrix contains the eigenvectors of A, S is a diagonal matrix contains the eigenvalues, U^{-1} is the inverse of U .

In general, given D random variable $x = (x_1, x_2, \dots, x_D)^T$, finding a low dimension of x , $s = (s_1, s_2, \dots, s_d)^T$ such that $D \geq d$ and captures most of information in x . However, dimensionality reduction implies loss in data; the task here is to preserve as much information as possible, and determine the best d value. This value can be determined by the largest eigenvalues that are associated with eigenvectors of the covariance matrix.

Mathematically PCA can be represented as:

Given a set of N data point $x_i \in \mathbb{R}^D$

Finding the first direction (unit vector) u such that $\sum_{i=1}^N (u^T x_i)^2$ is maximized.

Introduce Lagrange multiplier to enforce $\|u\| = 1$ and find a stationary point of:

$$L(u, \lambda) = \sum_{i=1}^N (u^T \cdot x_i)^2 + \lambda(1 - u^T u) \quad (3.3)$$

w. r. t u

$$L(u, \lambda) = \sum_{i=1}^N u^T (x_i x_i^T) u + \lambda(1 - u^T u) = u^T S u + \lambda(1 - u^T u) \quad (3.4)$$

where S is the covariance matrix

$$\frac{\partial L(u, \lambda)}{\partial u} = 2Su - 2\lambda u \quad (3.5)$$

$$\text{Set } \frac{\partial L(u, \lambda)}{\partial u} = 0$$

$$2\mathbf{S}\mathbf{u} = 2\lambda\mathbf{u} \quad (3.6)$$

\mathbf{U} is the Eigenvector of matrix S . thus the variation is maximized by the Eigenvector \mathbf{u}_1 corresponding to the λ_1 Eigenvalue of S . \mathbf{u}_1 is the first principle component. The second principle component is chosen such that it has the most variance and is orthogonal to the first one.

It can be seen after continuing in this manner that the d principle component of the data is the d Eigenvector of S with d largest Eigenvalue.

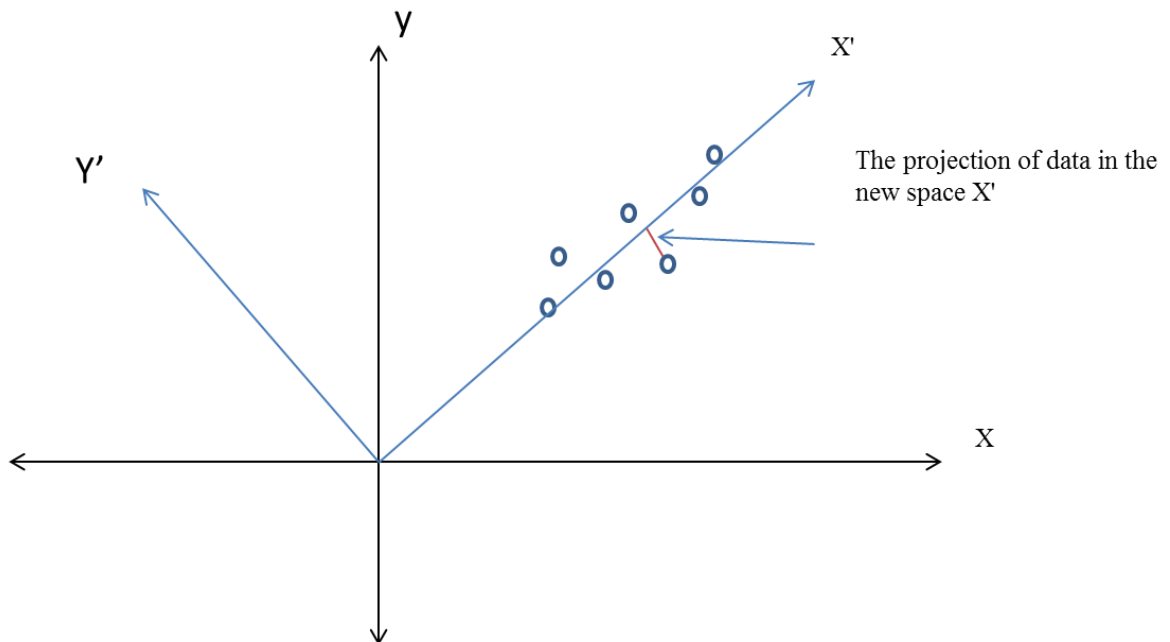


Figure 3-4: PCA Dimension reduction

Figure (3-4) illustrates how PCA is used in dimensionality reduction. The data in the Figure have two dimensions that are reduced to one dimension. In this case, PCA finds a lower dimension X' (one dimension) that the data is projected in with a smallest error.

The typical PCA algorithm can be applied as follows:

1. Given a set of N data points $\mathbf{x}_i \in \mathbb{R}^D$.
2. Subtract the mean from the data. $\mathbf{x}_i \leftarrow \mathbf{x}_i - \mathbf{c}$, where $\mathbf{c} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$
3. Find the covariance matrix. $\mathbf{S} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T$
4. Perform eigenvalue decomposition to construct the eigenvectors.

$$\mathbf{S} = \mathbf{U} \mathbf{D} \mathbf{U}^T \quad (3.7)$$

Then the *principal components* are the columns u_i of U ordered by the magnitude of the eigenvalues.

5. Project the sample to the new space.

$$\mathbf{y} = \mathbf{U}_d^T \mathbf{x} \quad (3.8)$$

Where d is the rank, that is $d < D$ and \mathbf{y} is the feature vector.

Figure (3-5) presents the mean of the mammography images. In Figure (3-6) the top 70 basis of PCA are shown.



Figure 3-5: Mean image of mammograms.

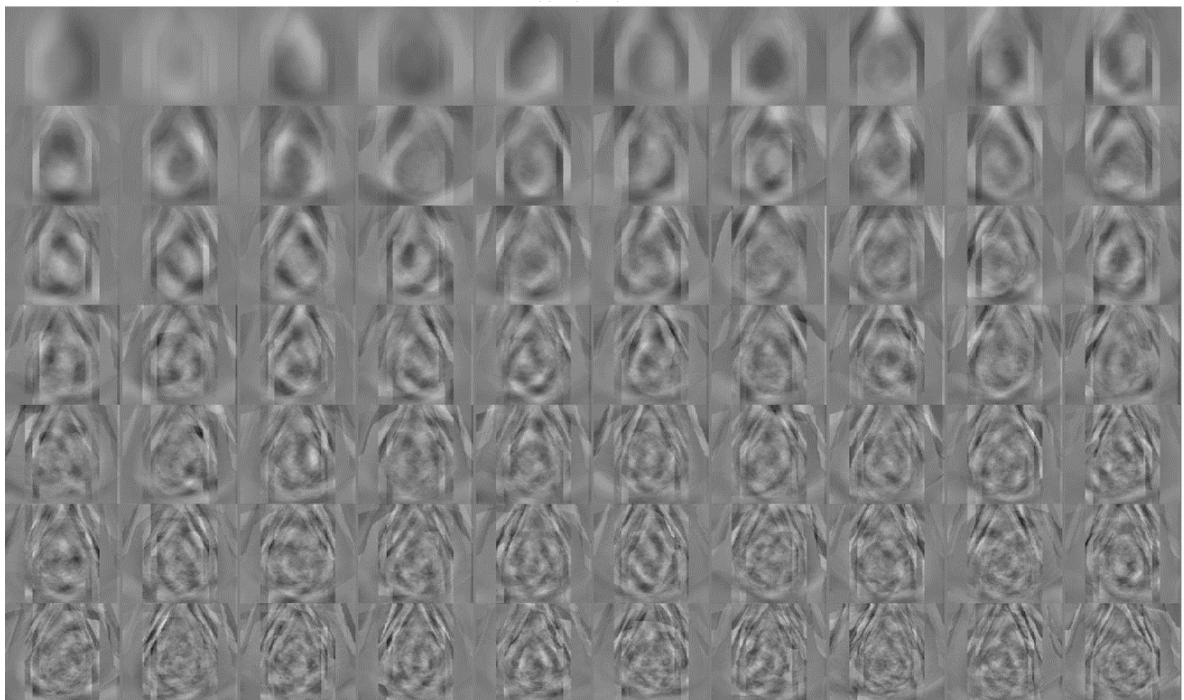


Figure 3-6: Top 70 PCA bases.

3.1.2.2 Two-Dimensional Principle Component Analysis (2DPCA)

Many techniques proposed in the literature use 2DPCA as feature for density classification. This technique is an extension to the standard PCA. It was developed by Yang et al. [54] to reduce the computations of the standard PCA.

Consider a matrix A of $m \times n$ and a vector x of N -dimensional unit vector. By projecting A onto x an M -dimensional vector y yields.

$$Ax = y \quad (3.9)$$

2DPCA finds a good projection vector x by using the total scatter of the projected samples. The trace of the covariance matrix of projected feature vector, which characterize the total scatter, can be obtained by maximizing this criterion:

$$J(x) = tr(S_x) \quad (3.10)$$

S_x is the covariance matrix written by

$$S_x = E[(A - EA)x][(A - EA)x]^T \quad (3.11)$$

Hence,

$$J(x) = tr(S_x) = x^T E[(A - EA)(A - EA)^T]x \quad (3.12)$$

From the criterion (3.10) the training image samples becomes

$$J(x) = x^T \left[\frac{1}{n} \sum_{i=1}^n (A(i) - \bar{A})^T (A(i) - \bar{A}) \right] x \quad (3.13)$$

Where \bar{A} is the average image of the training samples. This leads to the eigenvalue problem, where the eigenvector that maximize $J(x)$, of the covariance matrix correspond to the largest eigenvalue.

Thus, for a given image A , let $Y_k = AX_k$ where $k=1, 2 \dots d$. d corresponds to the largest eigenvalues and Y is the feature vector obtained from the projection.

3.1.2.3 Singular Value Decomposition (SVD)

Like PCA, SVD used to reduce the high dimension of the data and represent it using a lower number of dimensions, while maintaining as much information as possible.

In PCA, a symmetric matrix A can be factored into

$$A=UVU^T \quad (3.14)$$

Where U is an orthogonal matrix and V is a diagonal matrix of eigenvalues of A . For square matrix the decomposition of A can be

$$A=UVU^{-1} \quad (3.15)$$

Where U is an invertible matrix and V is a diagonal matrix.

If A is not symmetric or square, then such factorization is not possible. SVD can find a composition for any matrix. For dimensionality reduction of mammograms, number of methods were used to classify the density of the breast. Since the texture feature vector has high dimensionality, it is appropriate to choose a method that reduces the dimensionality and preserve the texture representation in a way that makes the system efficient. Singular Value Decomposition (SVD) is a good choice in this case. SVD is a matrix factorization technique that maps correlated variables into uncorrelated set.

$$A_{mn} = U_{mm}S_{mn}V_{nn}^T \quad (3.16)$$

Where A is a matrix of m columns and n rows, the columns of U are an orthogonal eigenvectors of AA^T , the columns of V are eigenvalues of $A^T A$, and S is a diagonal matrix contains the square root of the eigenvalues of U or V in descent order (singular values).

SVD can be used to calculate PCA. In case of applying eigenvalue decomposition in PCA and U in equation (3.7) is singular, then SVD is used to compute the Eigenvectors of PCA.

3.1.2.4 Non-negative Matrix Factorization (NMF)

Data in real world come naturally in non-negative format. However, when factoring the data matrix, the estimated data factors should have meaning or physical sense. This can be done if the estimated factors are non-negative.

NMF, as the new feature introduced in this thesis, is an unsupervised learning approach that leads to parts based representation. This representation comes from the additive combinations of original data, unlike other factorization methods such as PCA, ICA that allow subtractive combinations and learn a holistic representation. This non-negative constrains makes the resulted factors from the decomposition have only non-negative entries.

The NMF problem can be stated as follows:

Given a matrix $A_{m \times n}$, NMF decomposes this matrix into two non-negative matrices:

$$A_{m \times n} \approx W_{m \times r} P_{r \times n} \quad (3.17)$$

Or

$$A_{m \times n} = W_{m \times r} P_{r \times n} + E_{m \times n} \quad (3.18)$$

The columns of \mathbf{W} are called NMF basis, the rows of \mathbf{P} are their encoding coefficients, \mathbf{E} is the estimated error matrix and the rank r is chosen such that $(n + m)r < n \times m$.

The difference between NMF and PCA is that the columns of \mathbf{W} in PCA are orthogonal and the rows of \mathbf{P} are orthogonal to each other. Moreover, the entries of \mathbf{W} and \mathbf{P} are signed which makes the interpretation of basis have no intuitive meaning. NMF on the other hand, does not allow negative entries in \mathbf{W} and \mathbf{P} .

The NMF technique has been used in many applications including face recognition, gene expression, music analysis and others.

The basis of mammograms obtained using NMF are shown in Figure (3-7). Furthermore, these basis in this figure look like new images - real mammograms - and this makes the interpretation easier than PCA.

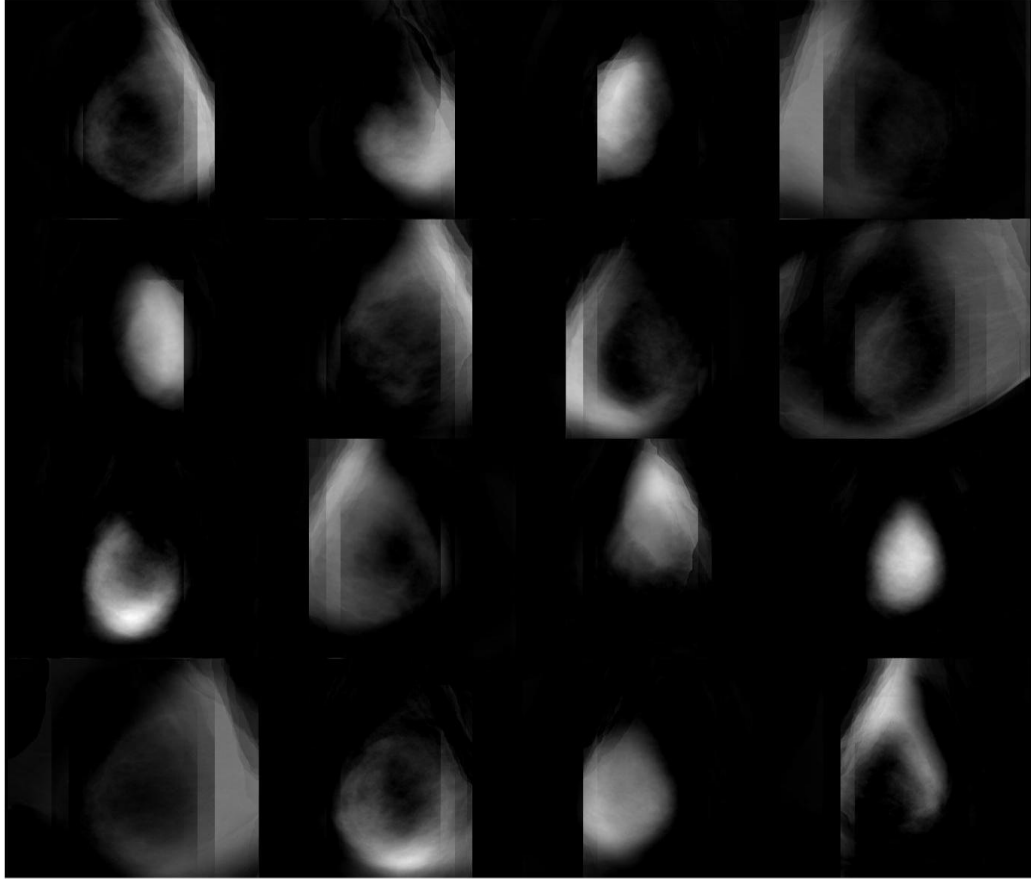


Figure 3-7: NMF bases.

3.1.2.5 NMF Algorithms

Many algorithms have been proposed to solve NMF problem are provided in [55][56].

These algorithms and their optimizations find the best possible solution from a set of feasible solutions that solve an objective function.

Lee and Seung [57] suggested, to create the NMF of the data matrix A , an approach that iteratively update the factors based on objective function. They choose the following objective function:

Given a nonnegative matrix $A \in R^{m \times n}$, find $W \in R^{m \times k}$ and $H \in R^{k \times n}$ such that $W > 0$, $H > 0$ and $k < \min(m, n)$ to minimize the functional

$$f(W, H) = \|A - WH\|_F^2 = \sum_{ij} (A_{ij} - (WH)_{ij})^2 \quad (3.19)$$

This Forbenius norm is used to measure the error between matrix A and its approximation WH . To achieve the desired objective function, the multiplicative update rules are used.

The multiplicative update rules used to solve Forbenius norm are:

$$W_{ij} \leftarrow \frac{(AH^T)_{ij}}{(WHH^T)_{ij}} W_{ij} \quad (3.20)$$

$$H_{ij} \leftarrow \frac{(W^T A)_{ij}}{(W^T WH)_{ij}} H_{ij} \quad (3.21)$$

Another commonly used objective function, called Kullback-Leibler divergence objective function:

$$D(A \parallel WH) = \sum_{ij} (A_{ij} \log \frac{A_{ij}}{(WH)_{ij}} - A_{ij} + (WH)_{ij}) \quad (3.22)$$

This measure is not a distance measure like the previous one, instead it measures the information about how the A probability distribution is close to the model distribution. To speed the convergence, the following update rules are commonly used:

$$W_{ij} \leftarrow \frac{\sum_{k=1}^N A_{ik} / (WH)_{ik} h_{ik}}{\sum_{k=1}^N h_{ik}} W_{ij} \quad (3.23)$$

$$h_{jk} \leftarrow \frac{\sum_{i=1}^M A_{ik} / (WH)_{ik} w_{ij}}{\sum_{i=1}^M w_{ij}} h_{ik} \quad (3.24)$$

Itakura-Saito [58] objective function can be expressed as:

$$D(A \parallel WH) = \sum_{ij} \left(\frac{A_{ij}}{(WH)_{ij}} - \log \frac{A_{ij}}{(WH)_{ij}} - 1 \right) \quad (3.25)$$

This divergence was obtained from the maximum likelihood (ML) estimation in a short time speech spectra as a goodness measure between two spectra to fit.

3.1.2.6 NMF vs. PCA

PCA has many properties and advantages. PCA can produce the best rank approximation in the sense that $\|A - A_k\|$ is small. So A_k is the best approximation among all possible rank-K approximations. This is the optimality property. The algorithms used in computing the principle components are robust and accurate. This robustness and uniqueness are useful properties. Another property of PCA is that the basis vectors are orthogonal. However, the result basis vectors are mixed in sign which makes the interpretation of these vectors more difficult. Also, these vectors are completely dense, which require more storage. Although A_k provides best approximation, determining best truncated value k is hard and problem dependent.

NMF on the other hand, has different properties that eliminate these weaknesses of PCA. First the basis vectors are positive. This positivity constrain leads to ‘parts based representation’ which makes the interpretation of these basis possible. Another useful property is that the resulted factors are sparse. This leads to an efficient storage of these factors.

As shown, the strengths of PCA become the weaknesses of the NMF and vice versa. Unlike PCA in the unique factorization, NMF has no unique factorization. This convergence problem is the most important weakness issue of NMF. However, different

algorithms used to calculate NMF can produce different results and converge to dissimilar local minima, thus the initialization stage in these algorithms is critical.

3.1.2.7 Mean Density Feature

The idea of this feature stems from the fact that the tissue in dense mammograms has a high intensity value, while the background has low intensity value. If two thresholds could be found, then these thresholds, that separate tissue and background, can determine the amount of dense and fat in the whole breast.

The features are simply extracted by calculating the amount of fat and tissue (density) to the total area of the breast. The ratio of fat is determined by

$$fat_{ratio} = \frac{fat}{fat + dense} \quad (3.26)$$

On the other hand, the ratio of density measured using:

$$dense_{ratio} = \frac{dense}{fat + dense} \quad (3.27)$$

These two ratios are used as features to determine the breast density.

To determine the value of the thresholds the histogram of the density types is used. Figure (3-8) illustrates an example of dense image histogram where the first red line represents the values of the background of the image and between the two red lines the fat values are fit while the values after the second red line represent the dense.

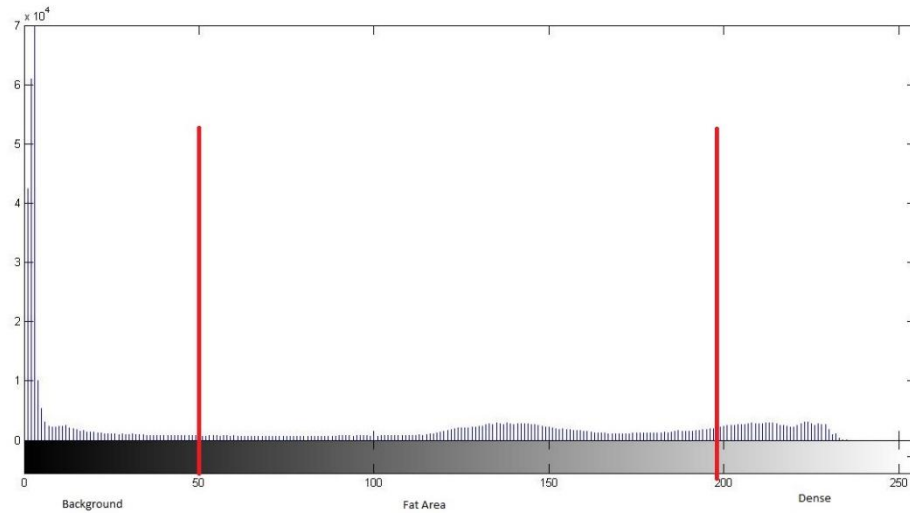


Figure 3-8: histogram of three densities, (Left) fatty glandular, (Middle) fatty, (Right) dense

It can be seen from this figure that the background takes the grey level value of less than 50, while tissue takes more than 180. These values can vary from one mammogram to another. In the experiment, different values have been tested, but these values produce the best results.

3.1.2.8 Local Binary Pattern (LBP) Feature

LBP is a widely used technique by researchers due to its simplicity and efficiency. This technique is one of the texture analysis techniques used combined with other features in many approaches in the literature to model the density of the mammogram.

The local texture information is encoded into a binary value. This is done by analysing the joint distribution of grey level of the neighbouring pixels in a local neighbourhood [40]. The value of grey level of the centre point is subtracted from the neighbour pixels. If the difference is less than zero, then a value of 0 is assigned to the neighbour pixel and 1 otherwise. The binary label assigned to the neighbour pixels are transformed into a unique LBP number by:

$$LBP = 1 + \sum_{p=0}^{p-1} s(g_p - g_c)2^p \quad (3.28)$$

Where g_c is the grey level value of the center pixel, g_p represent the grey level value of the neighboring pixel, $s(x) = 1$ if $x \geq 0$ else $s(x) = 0$ the histogram of the LBP is populated to represent the texture image.

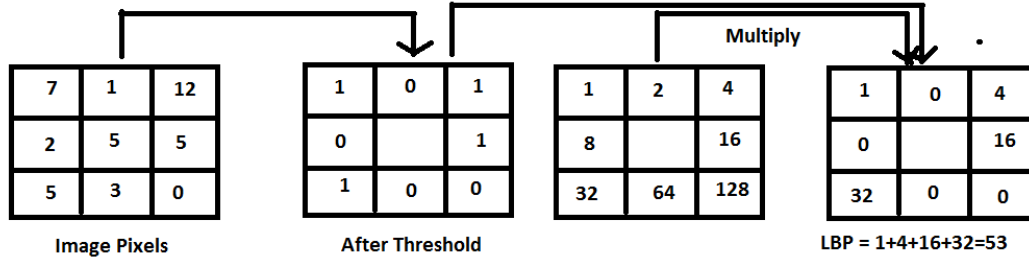


Figure 3-9 calculating LBP for pixel

An example that demonstrates the LBP calculation is shown in Figure (3-9). In this example to calculate the LBP for the middle pixel, the neighbour pixels values are compared to the value of the middle one. If these values are greater or equal to the centre value, then 1 is assigned to the output matrix, otherwise 0 is assigned. This output matrix is multiplied by a matrix where it is values are to the power of 2. The LBP value of the middle pixel is the sum of the output matrix values.

3.1.3 Classification

In the classification step, the classification algorithm is provided with training samples in which it can learn to predict new coming samples. This step is called the supervised learning step. Moreover, The supervised learning approximates a function that maps a vector into given classes by examining the training examples. In this step several tests

have been done to find best parameters that improve the output of the system and increase the result accuracy.

3.1.3.1 Support Vector Machine (SVM)

To retrieve the images that are related to a specific class based on BI-RADS, Support Vector Machine (SVM) is a good choice for the binary classification. SVM is a learning technique which has been applied to CBIR and shows promising results [59].

The SVM classifier finds the linear decision boundary - hyper plane - that successfully separates the points of a given two classes and a group of points that belong to these classes. This hyper plane maximizes the distance between the two classes. It can be determined by a set of points called support vectors. In practice instead of using a straight line in separation, a function is used called kernel. Polynomial and Gaussian kernels are different types of kernels that are widely used.

Figure (3-10) shows the architecture of the linear SVM. The greater the distance between the classes, the better the classification.

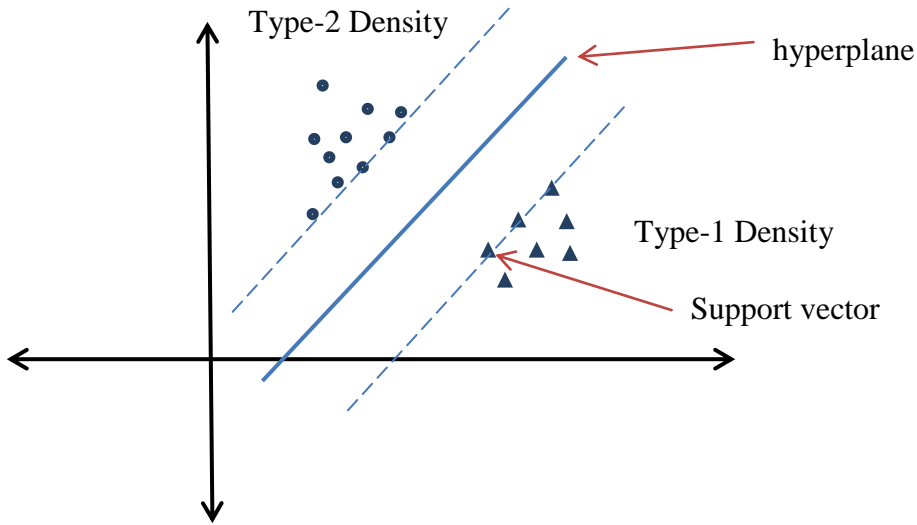


Figure 3-10: linear SVM structure

SVM works in two stages: The learning stage in which it learn a decision boundary for the training data set and the classification stage in which it predicts the class of the new input based on the learned decision boundary.

Given a set of training data

$$D = \{(x, y) | x \in R^p, y \in \{1, -1\}\}$$

Y is the class labels, indicates that x belongs to either +1 or -1 classes.

In the simplest form, SVM are hyperplanes that best isolate the data with maximal margin. Suppose plus plane $w \cdot x + b = 1$ and the minus plane $w \cdot x + b = -1$

So the vector x is classified as +1 if

$$w \cdot x + b \geq 1$$

Or it is classified as -1 if

$$w \cdot x + b \leq -1$$

Where

$$-1 < w \cdot x + b < 1.$$

The margin width M is given by:

$$M = \frac{2}{\sqrt{w \cdot w}}$$

Then finding the largest margin while minimizing the classification error is solved as a convex quadratic programming problem. This gives the Lagrangian:

$$L_D = \sum_i a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j x_i x_j \quad (3.29)$$

Where

L_D is maximized with respect to a

a_i is Lagrange multiplier

y_i is the vector class

x_i is the train vector

Subject to $0 < a_i < C$ where C is Constant chosen by user.

To classify vector x . the sign is calculated

$$f(x) = \sum_{i=1}^N a_i y_i s_i x + b \quad (3.30)$$

Where

N is the number of support vectors

s_i is support vector with class y_i

b can be calculated by choosing a support vector where $a < C$ with

$$b = \frac{1}{y_i} - \sum_{j=1}^N a_j y_j x_j x_i \quad (3.31)$$

The hyperplane is defined as

$$w \cdot x + b = 0 \quad (3.32)$$

Where

$$w = \sum_{i=1}^N a_i y_i s_i$$

SVM can be used with different kernel functions. Furthermore, kernel functions map the input vector to a higher dimensional space where a better hyperplane with minimal classification error can be found. The kernel functions are defined as

$$K(x_1, x_2) = \Phi(x_1) \cdot \Phi(x_2) \quad (3.33)$$

Where $\Phi(x)$ maps x to another Euclidean space

The dot product $x_1 \cdot x_2$ can be replaced with $K(x_1, x_2)$ so the Lagrangian becomes

$$L_D = \sum_i a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j K(x_i, x_j) \quad (3.34)$$

The classification function become

$$f(x) = \sum_{i=1}^N a_i y_i K(s_i, x) + b \quad (3.35)$$

Where

$$b = \frac{1}{y_i} - \sum_{j=1}^N a_j y_j K(x_j, x_i) \quad (3.36)$$

The most common functions used by SVM are

Polynomial of degree p $K(x, y) = (x \cdot y + 1)^p \quad (3.37)$

$$\text{Radial basis} \quad K(x, y) = e^{-|x-y|^2 / 2\sigma^2} \quad (3.38)$$

3.1.3.2 K-Nearest Neighbor (KNN)

The KNN algorithm is non-parametric, lazy learning algorithm. This means that the data can have any structure and the training phase is very minimal. It classifies the object based on the majority votes of it is 'K' neighbors. The number of neighbors (K) influences the classification of the algorithm.

The KNN assumes that the training samples are in the feature space, each has a label. All the samples are stored in the training phase. In the classification phase, KNN identifies K neighbours, where K is a user defined variable, regardless of the labels. The test unlabelled samples are classified by the most K labels of their neighbours.

Figure (3-11) shows an example of classifying the green point with KNN.

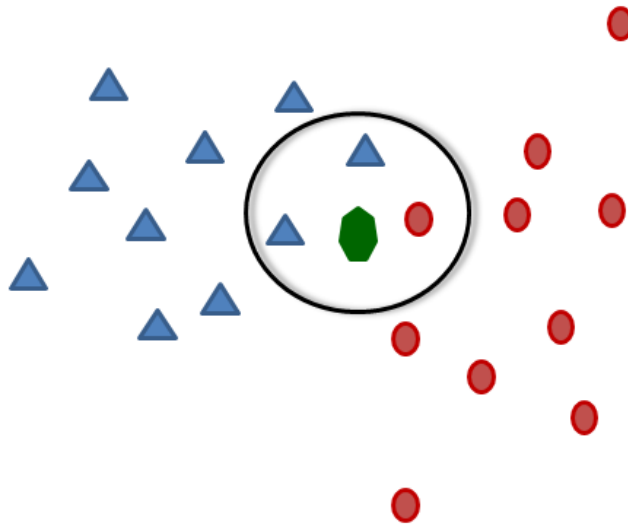


Figure 3-11: Classification using KNN (K=3).

CHAPTER 4

EXPERIMENT RESULTS

In this section, a description about the tools and the database used in this experiment is introduced first. Then the conducted experiment results on mammogram density classification are explained. Finally, in the summary section, conclusions about the features used and classifier performance are discussed.

4.1 Databases and Tools

Before discussing the performance and the results of the proposed system, a description of the tools and database is presented. However, several mammography databases are available for scientific research. In this thesis MIAS database is used, since it provides information about the density of the breast and it is commonly used by most of the approaches in the literature, so that a fair comparison between the proposed approach and other approaches can be possible.

4.1.1 Mammographic Image Analysis Society (MIAS) Database

This database is used in the experiment of the classification system. The database is freely available for scientific research. Moreover, the database images have been generated by a research group from UK and images were taken from UK National Breast Screening Programme and digitized using Joyce-Loebl scanning microdensitometer. The

database contains 322 images (50 micron pixel edge) in MLO view, but it has been reduced as 1024 x 1024 pixel images. The information that describes the density is available, however it includes also class of masses normality, severity type and the location of the abnormality.

The density of each image is annotated by expert radiologist and classified into three distinct categories: Fatty (F) contains 106 images, Fatty-Glandular (G) contains 104 images and Dense-Glandular (D) contains 112 images.

In the experiment, for binary classification the fatty and dense categories are considered, so a subset of this database, about 228 images, are used.

4.1.2 Development Tools

A set of tools and software applications that contributed in completing this work are listed:

- MATLAB: stands for MATrix LABoratory software, is high-level, prototype language used for numerical computation, data analysis and visualization. The proposed system was implemented using this software with an image processing toolbox.

- LS-SVM [60]: Least Squares SVM is a fast and free library available under GNU general public license policy.

- NMF-Tool [61]: is a free toolbox for research and non-profit application that contains a set of NMF algorithm implementations.

4.1.3 Classification

In the process of density classification, two well-known classifiers are used. 1) K-Nearest Neighbor (KNN). 2) Support Vector Machine (SVM). Each are tested with different parameters configuration to achieve higher performance.

KNN is one of the simplest and a powerful classifier. It classifies an instance by identifying K neighbors regardless of labels. The final decision is based on the most labels of the K neighbors.

SVM, on the other hand, is a very popular classifier. The kernel trick in SVM, provides a powerful way for nonlinear class separation. However, to find the optimal hyper plane between the training data sets, it requires long computations and time. To reduce these computations, LS-SVM is used.

4.1.4 Validation and Classification Performance

The proposed method can be validated using different approaches. In this experiment, K-Fold cross validation scheme is chosen with 10 folds.

The K-Fold cross validation model, divides the training data into k data sets. One of these subsets is used for testing while others are used for training. Then, another different set is chosen for testing each time and the average error rate is computed among all data sets.

The advantage of this model is that each data instance is used in testing exactly once, and in training $k-1$ times.

In machine learning, the evaluation can determine the usefulness of the classifier on various datasets, since no single classifier can provide best results on all problems. Based on confusion matrix, the following evaluation measures have been used:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.1)$$

$$Recall = Sensitivity = TPR = \frac{TP}{FN + TP} \quad (4.2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4.3)$$

The accuracy of a classifier is the correct prediction over all samples. In the same context regarding classifiers performance investigation, sensitivity and specificity concepts are also introduced. Sensitivity is the amount of the first class (fatty in our case) that predicted as fatty. Specificity on the other hand, is the amount of the second class (dense in our case) that predicted as dense. Basically, the classifier should be sensitive and specific as much as possible.

In addition to these metrics, Area under Receiver Operating Characteristics (ROC) is also considered. In order to decide which classifier is better than the other, the ROC performance is reduced to a scalar value that represents the expected performance. A common approach is to calculate the area under ROC (AUC), where the total area is equal to one. Larger area implies better performance of the classifier [62].

4.2 Results

Six different feature types have been used to identify the mammogram density: PCA, 2D PCA, SVD, NMF, LBP and the Mean density features. After computing these features, the system saves these features in a database file. Moreover, the results obtained from each feature using the KNN and SVM classifiers are analysed and discussed in following section.

4.2.1 PCA Results:

In order to determine the number of components that can be used in the feature extraction step, the cumulative sum of Eigenvalues has been used. The amount of variance that each eigenvector represents can be calculated by dividing the eigenvalue that correspond to that eigenvector by the total sum of eigenvalues. Plotting of the cumulative sum of eigenvalues provides a good indication for the number of components that should be used in feature extraction. Figure (4-1) illustrates the cumulative sum of PCA eigenvalues on mammograms database. As shown in the Figure (4-1), only 47 components capture more than 96% of the data information, therefore instead of using the whole dimensions, only these 47 dimensions will be used.

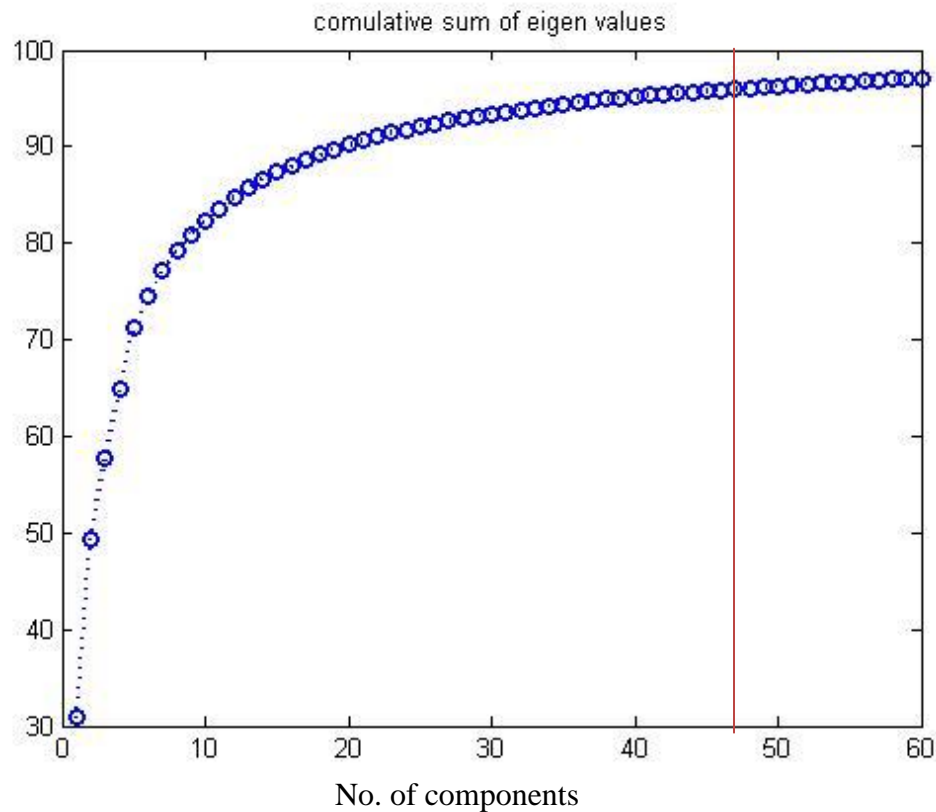


Figure 4-1: Cumulative sum of eigenvalues used in PCA decomposition.

In mammogram patches, the number of components used is less than the number of components used in mammograms database. This is clearly obvious in Figure (4-2), where only 34 components are considered.

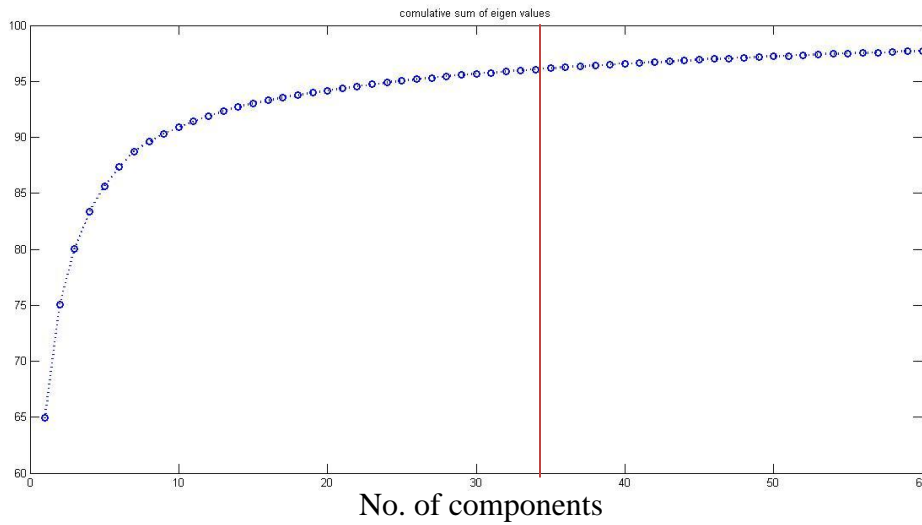


Figure 4-2: Cumulative sum of eigenvalues of mammogram patches used in PCA decomposition.

After determining the number of components, the experiment was conducted on several odd components from 1 to 45 on mammograms dataset and from 1 to 33 on patches dataset. Figures (4-3) and (4-4), illustrate the mammogram results using SVM with different kernels and using KNN with different K neighbours.

The horizontal axes show the number of components, while the vertical axes represent the performance measure.

From Figure (4-3-a) the SVM classifier with polynomial kernel achieves the best results (87%) utilizing 9 components. Moreover, with the same number of components, the RBF kernel reaches a close accuracy (86%). On the other hand, the best recognition rate of KNN classifier is around (84%) with 11 components when K is set to 5 which is slightly lower than SVM. Other performance measures, Figures (4-3-b) (4-3-c) (4-3-d), provide high results. Also, Figures (4-5) and (4-6) shows the results of both classifiers on patches dataset.

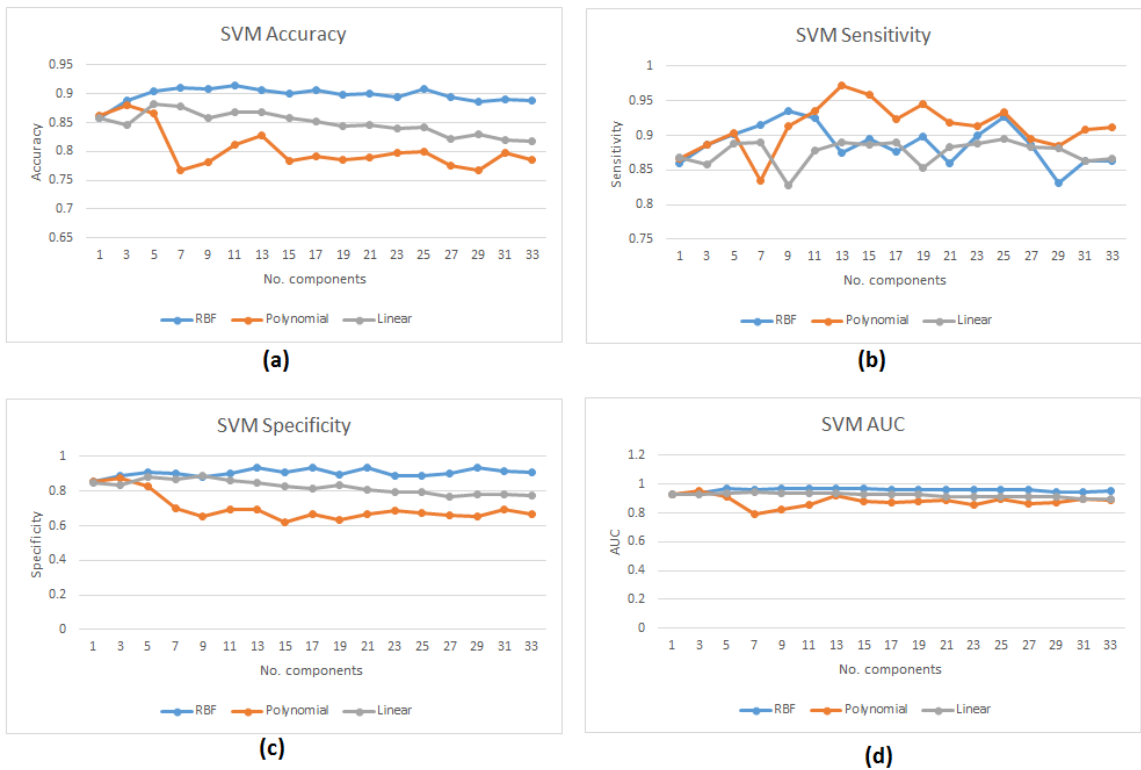


Figure 4-5: PCA results using mammogram patches and the SVM method.

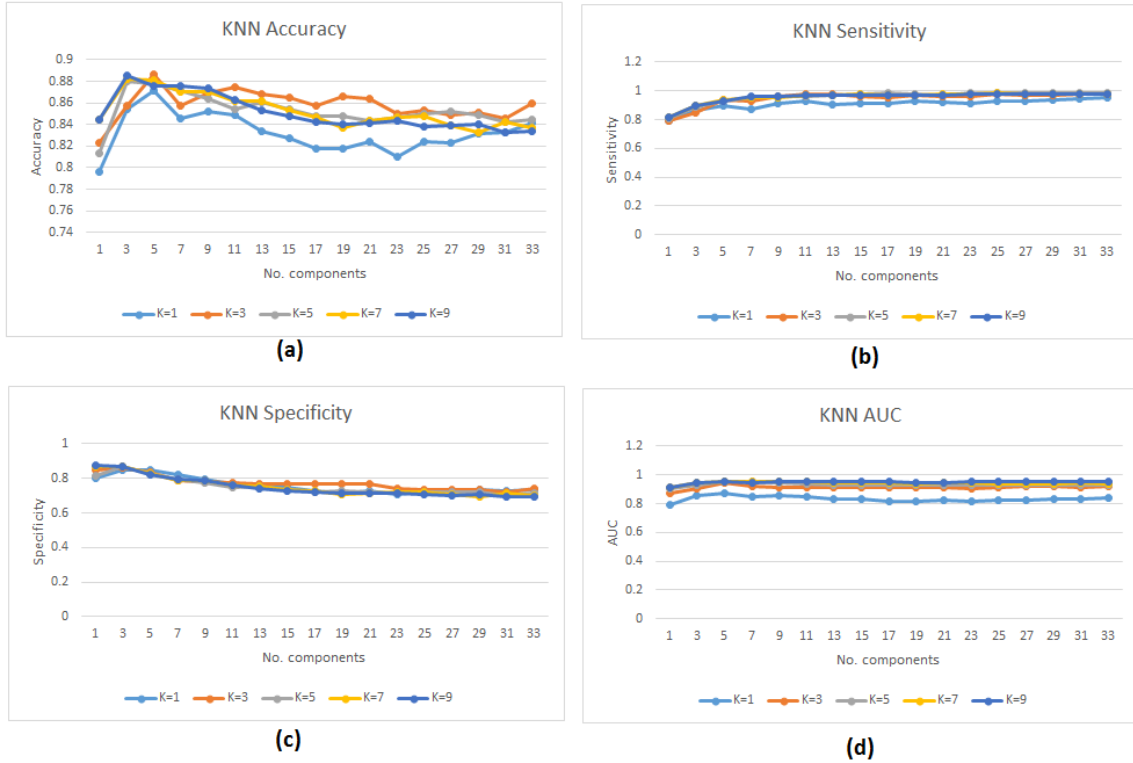


Figure 4-6: PCA results using mammogram patches and the KNN method.

From the above Figures (4-5-a) and (4-6-a), the SVM performance is improved and reaches up (91%) with RBF kernel by considering 11 components, whereas KNN classifier achieves recognition rate (89%) when K is set to 3 and 5 components are considered or K is set to 9 and 3 components are selected.

The best results are obtained on the first 11 components using SVM and KNN on both datasets. However, a significant improvement in the performance of both classifiers is achieved when patches dataset is used. Moreover, the SVM classifier outperforms KNN on both datasets. Although SVM with RBF kernel achieves the best results, the sensitivity of this kernel in patches dataset is fluctuating on different components as indicated in Figure (4-5-b).

The total time (in seconds) for extracting the PCA feature is presented in Table (4-1).

Table 4-1: Estimated time for feature extraction using the PCA method.

	Mammograms	Patches
Feature Extraction Time (Seconds)	429.7	408.2

The time for extracting PCA feature from patches is less than mammograms. So, using patches database improves performance and feature extraction time of the proposed system.

4.2.2 2D PCA Results

This feature is faster than the standard PCA. Since it follow the same PCA procedure, we need to determine the number of principle components in feature extraction. This is done by using the same technique in standard PCA. The cumulative sum of eigenvalues is shown in Figure (4-7) and the number of components used are 11.

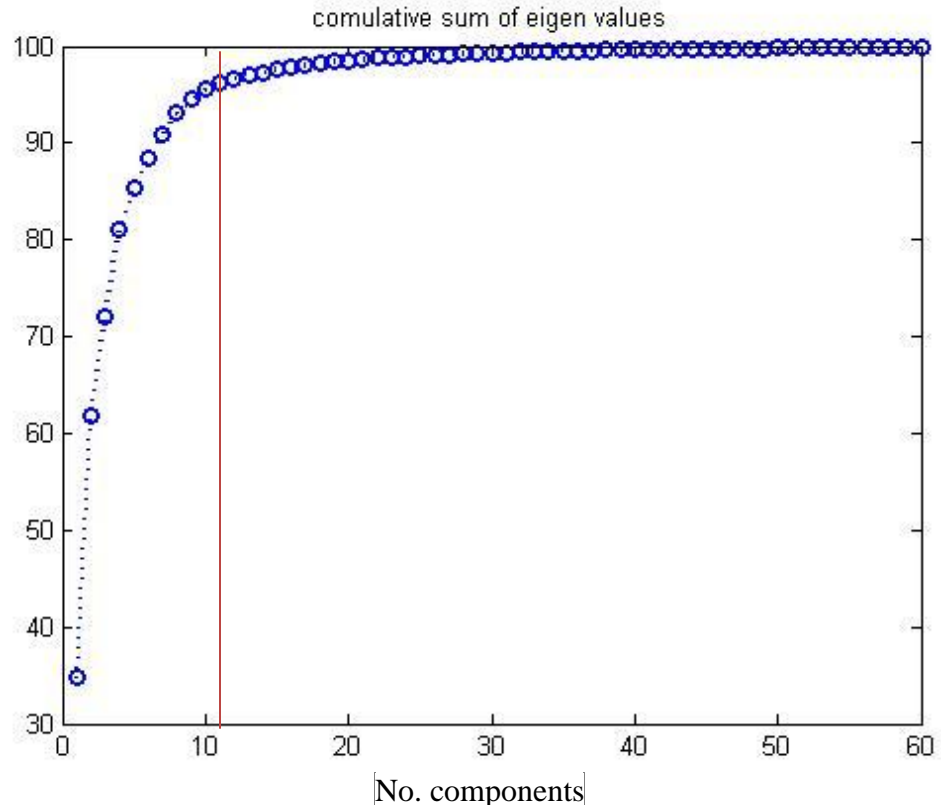


Figure 4-7: Cumulative sum of eigenvalues on mammograms used in 2DPCA decomposition.

Using mammogram patches, only 6 components are selected for further processing. Figure (4-8) shows the plot of cumulative sum of eigenvalues on the dataset of mammogram patches. To enhance the classification performance, the experiment is conducted using 10 components instead of 6.

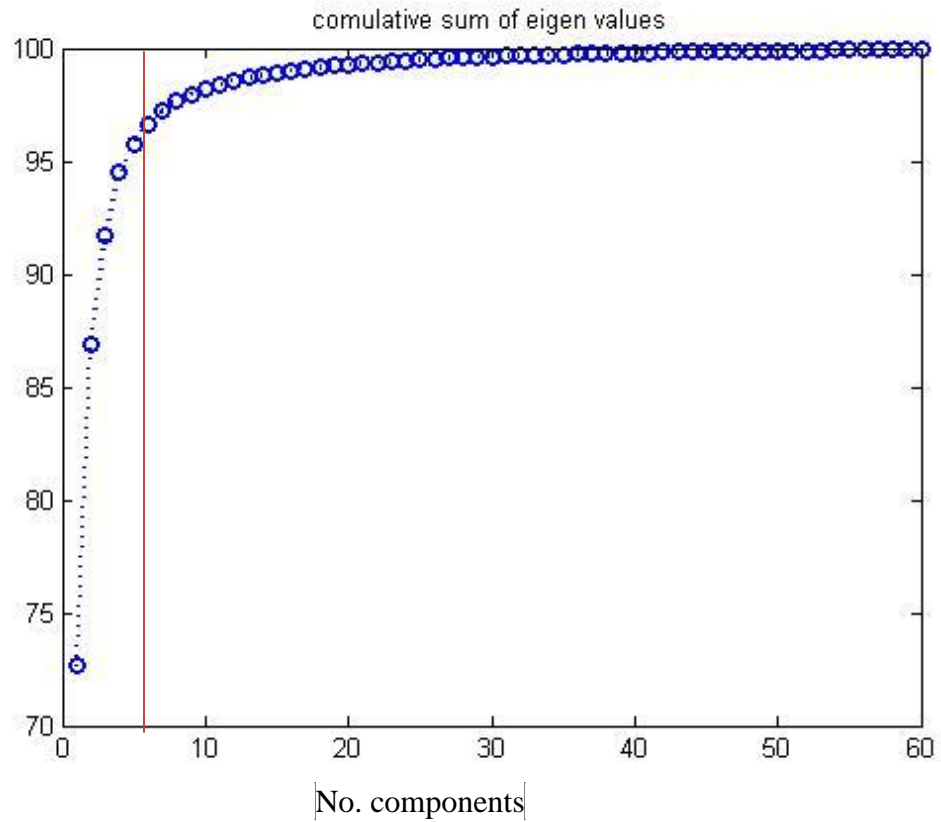


Figure 4-8: Cumulative sum of eigenvalues on patches used in 2DPCA decomposition.

The obtained results on mammograms are shown in Figure (4-9) for SVM with different kernels and Figure (4-10) for KNN with different K neighbours.

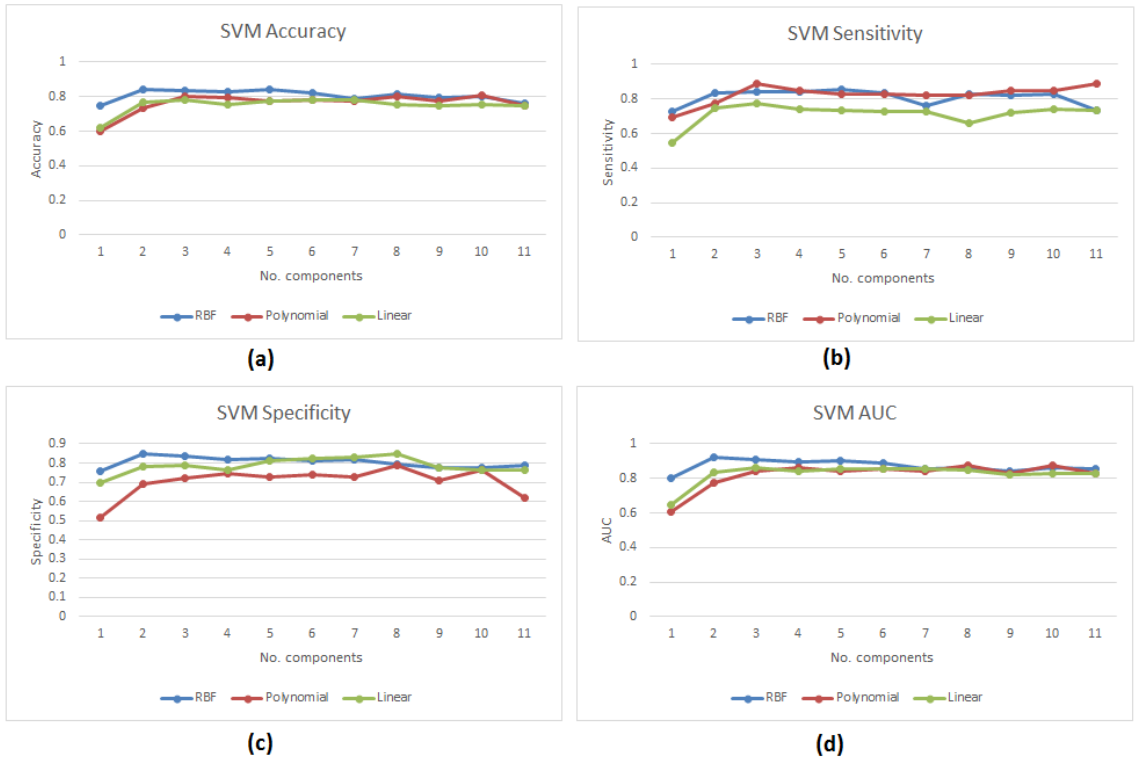


Figure 4-9: 2DPCA results using full mammograms and the SVM method.

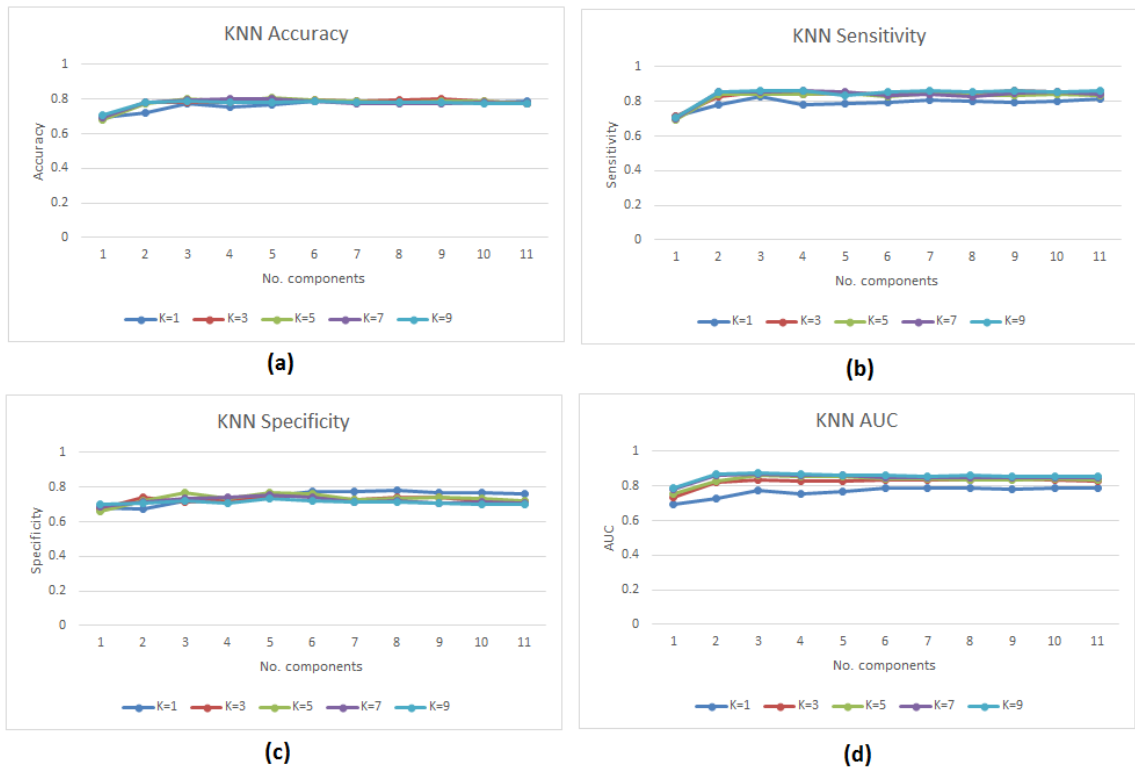


Figure 4-10: 2DPCA results using full mammograms and the KNN method.

The best results could be observed when using SVM with RBF kernel (85%) considering only 2 components. While, KNN achieves (80%) when K is 5 or 7 and 5 components are used.

Applying the same classifiers on mammogram patches, the results are illustrated in Figure (4-11) for SVM and Figure (4-12) for KNN.

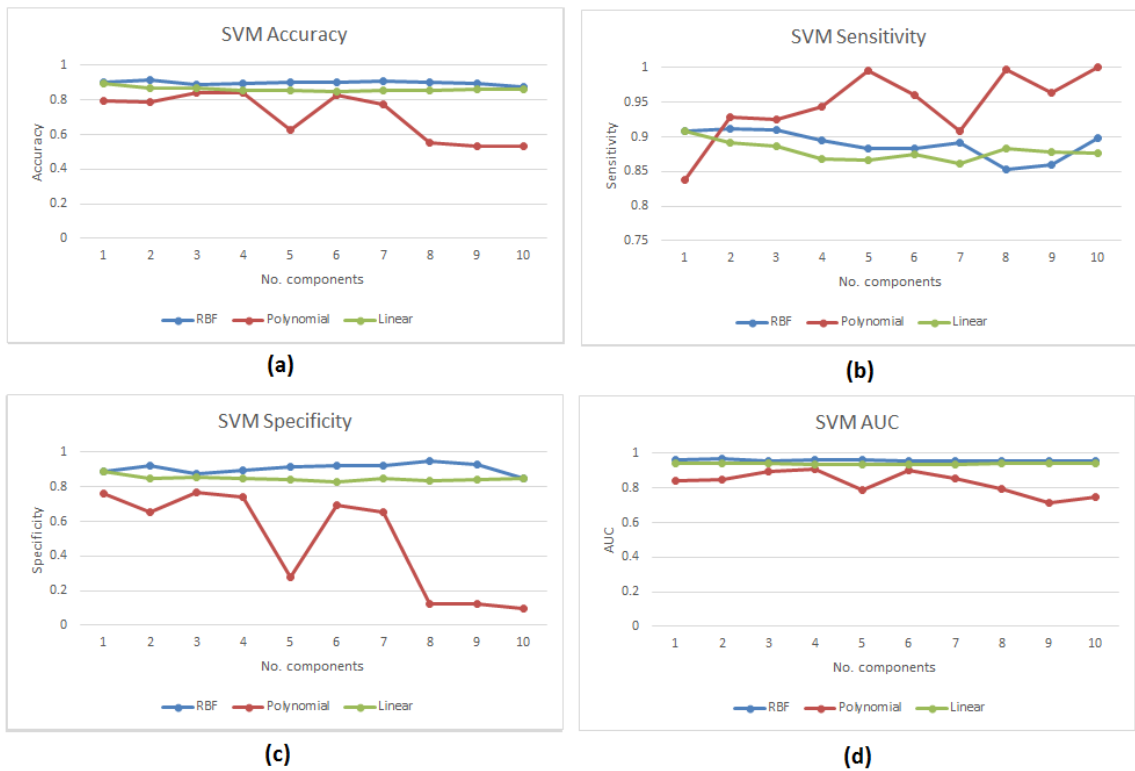


Figure 4-11: 2DPCA results using mammogram patches and the SVM method.

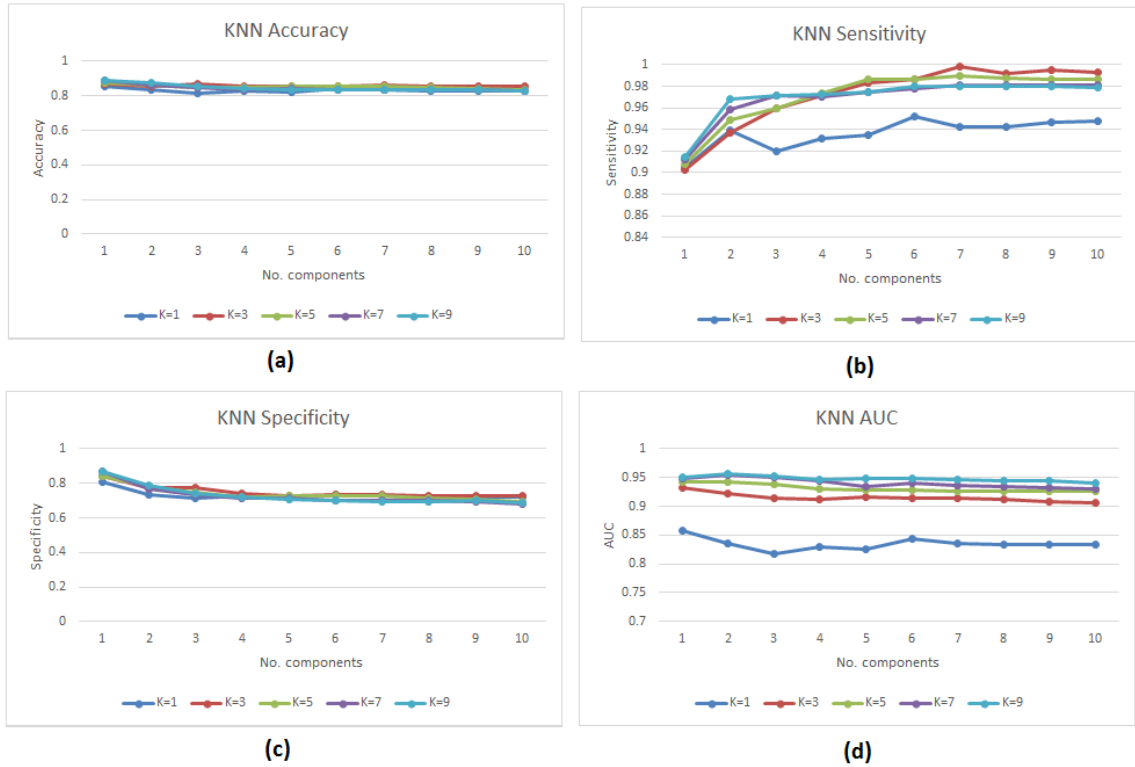


Figure 4-12: 2DPCA results using mammogram patches and the KNN method.

The performance of these classifiers increase when the patches are used. RBF kernel provide the highest accuracy (92%) with 2 components. Linear kernel provide close results (90%) with 1 component, while polynomial kernel reaches up (84%) with 3 components. KNN achieves (89%) recognition rate when K is 7 or 9 with 1 component.

The 2D PCA feature is much faster than standard PCA. Table (4-2) represents the total time for extracting 2DPCA feature.

Table 4-2: Estimated time for 2D PCA extraction.

	Mammograms	patches
Time (seconds)	0.75	0.29

4.2.3 SVD Results

The SVD technique is more stable than the standard PCA. It can capture more information in a few basis (components). The number of basis used in mammograms are 5 demonstrated in Figure (4-13).

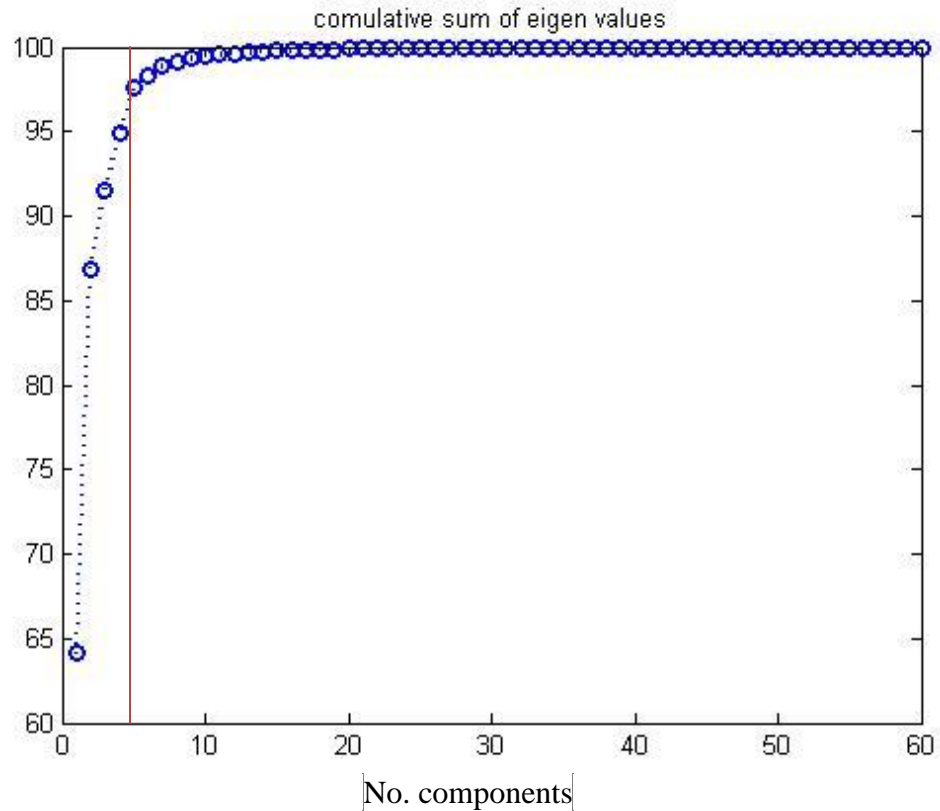


Figure 4-13: Cumulative sum of eigenvalues in full mammogram factorization using the SVD method.

Since SVD can capture most of information in one component on patches dataset as indicated in Figure (4-14). It should be noted that 10 components are retained for comparison purposes using both datasets (full mammograms and patches).

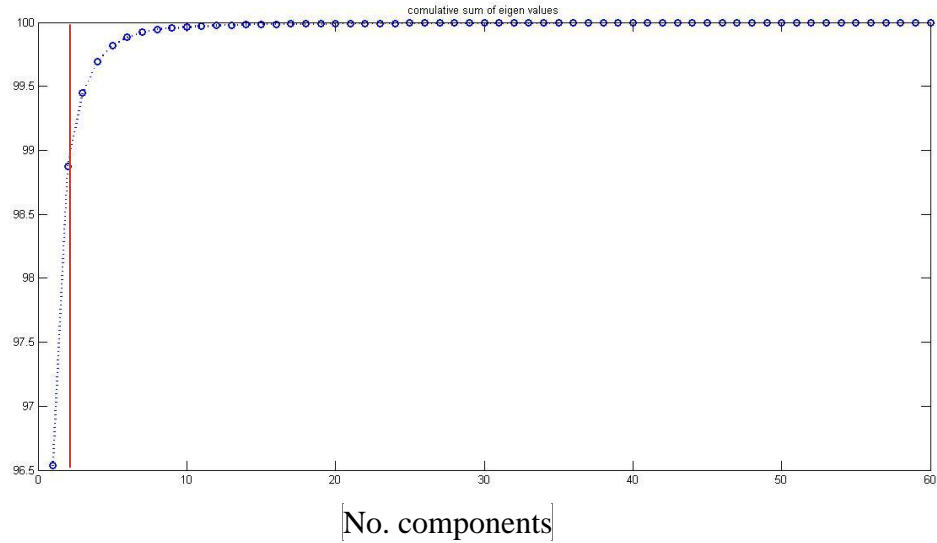


Figure 4-14: Cumulative sum of eigenvalues in mammogram patches factorization using the SVD method.

In comparison between SVM and KNN, the SVM classifier achieves better results than KNN. Figure (4-15) depicts the results of SVM and Figure (4-16) illustrates the results obtained using KNN on mammograms.

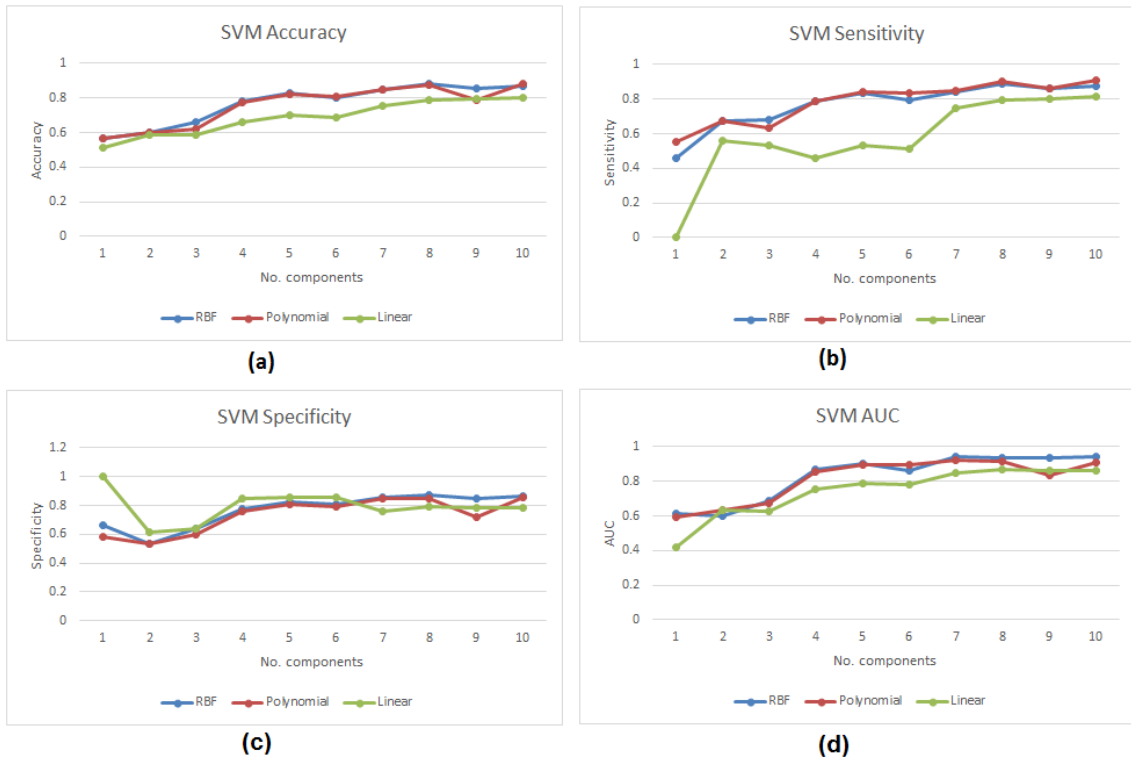


Figure 4-15: SVD results using full mammograms and the SVM method.

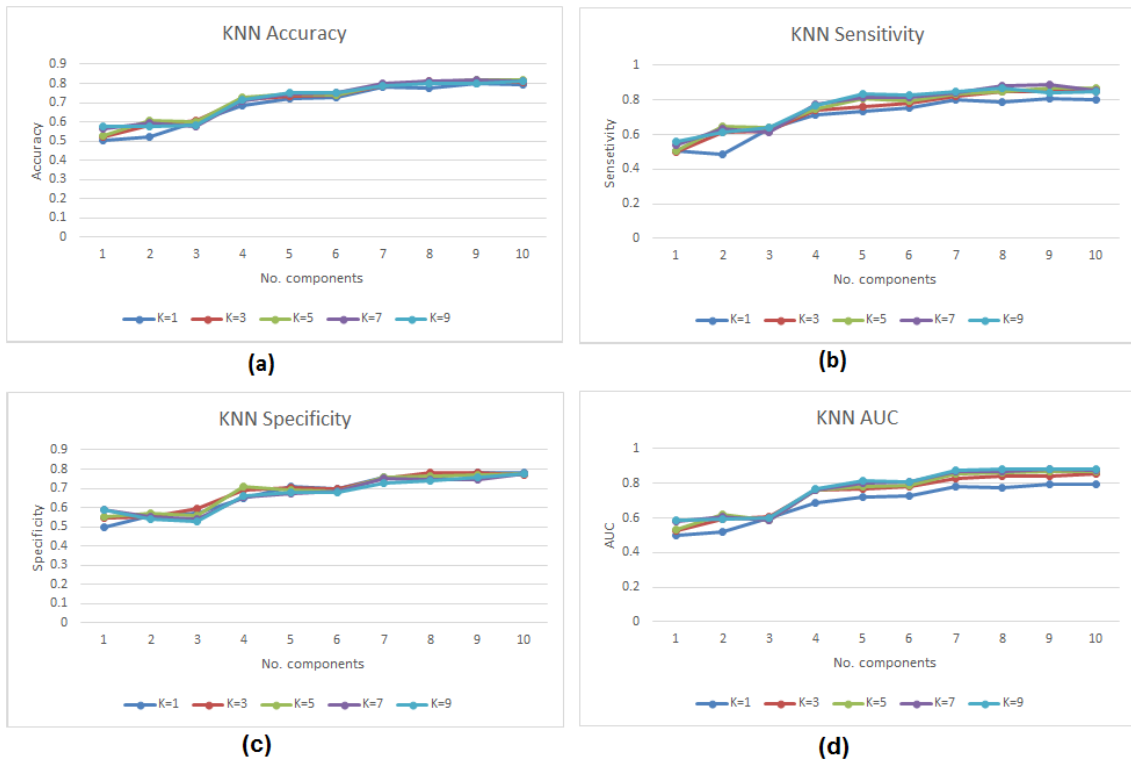


Figure 4-16: SVD results using full mammograms and the KNN method.

It is obvious that SVM with RBF kernel achieves best result (88%) with 8 components. The Polynomial kernel also has close results compared to the RBF kernel. Unlike SVM, KNN classifier reaches up to (82%) with 10 components and K set to 5.

Conducting the experiment on mammogram patches, the results for SVM and KNN are shown in Figures (4-17) and (4-18) respectively.



Figure 4-17: SVD results using mammogram patches and the SVM method.

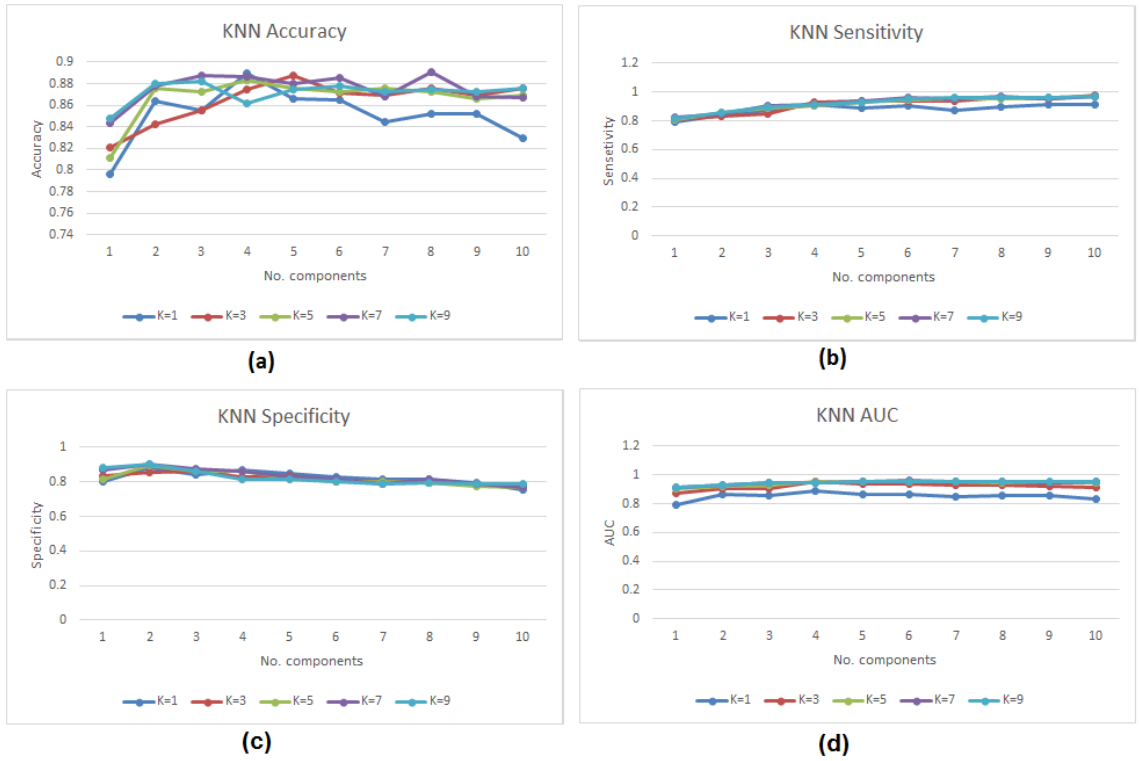


Figure 4-18: SVD results using mammogram patches and the KNN method.

With 8 components both classifiers achieve their best results. However, SVM still provide better results (91% with RBF kernel) than KNN (89% with K is set 7). But the overall performance of both classifiers on patches is better than mammograms.

SVD is slower than PCA in case of time required for feature extraction, this can be clearly shown in Table (4-3).

Table 4-3: Estimated time for SVD extraction.

	Mammograms	patches
Time (seconds)	895.99	820.37

4.2.4 NMF Results

NMF factorization works by factoring the matrix into two positive matrices, W which is called the basis and H which is the coefficients or features. Unlike PCA and SVD, NMF does not provide information above information capturing, so Peak Signal-to-Noise Ratio (PSNR) similarity measure is used. If the value of similarity measure between the training data matrix and the reconstructed matrix is big, then these matrices are more similar to each other's. The PSNR can be used to show how much information has been captured between the original image and the constructed image from the training steps. Since the quality between two images is calculated, the higher resulted value the higher the quality. However, the typical value of PSNR for 8-bit images is 30 dB . Therefore, this value is used to determine the number of basis used in factoring the data matrix using NMF. On mammograms dataset the number of basis used are 100, which is determined by the PSNR measure.

The obtained results are shown in Figure (4-19) for SVM and Figure (4-20) for KNN.

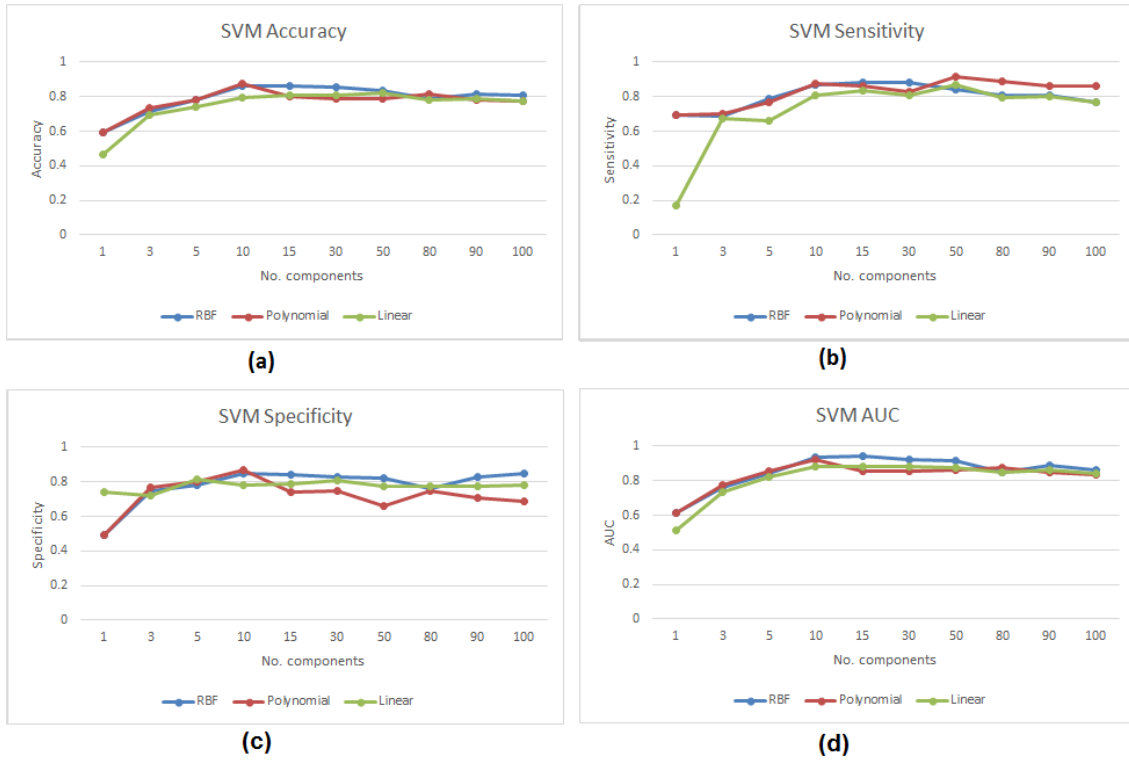


Figure 4-19: NMF results using full mammograms and the SVM method.

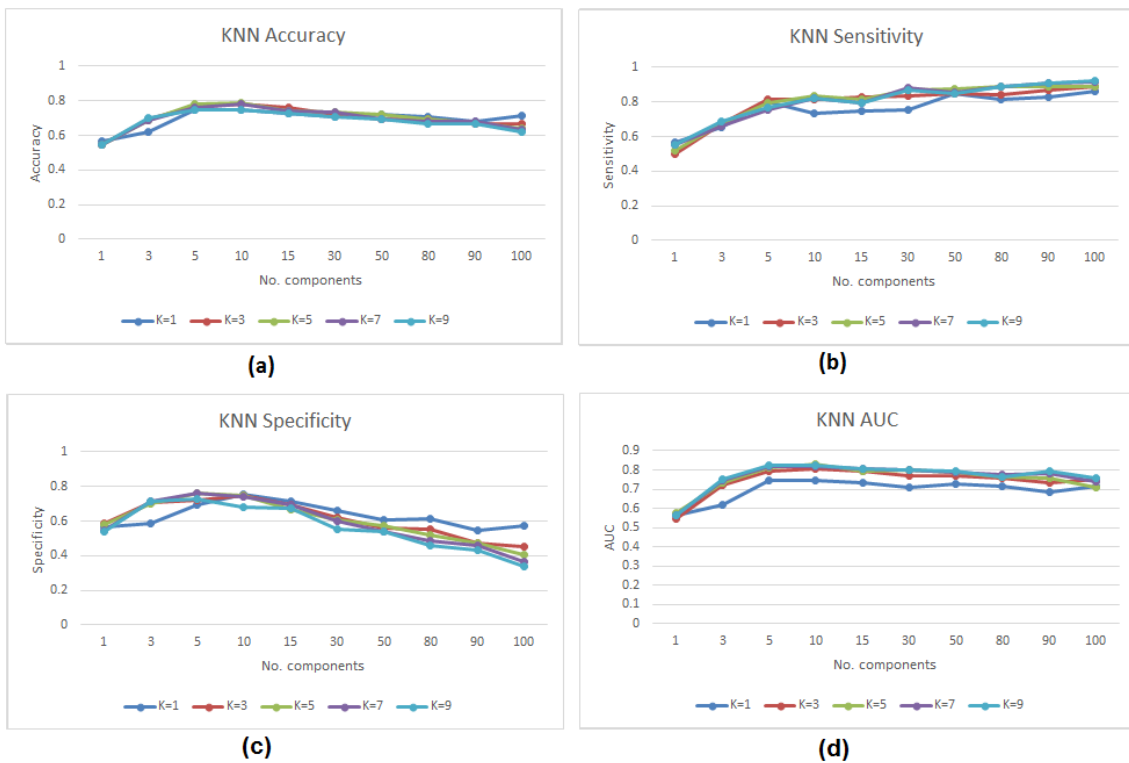


Figure 4-20: NMF results using full mammograms and the KNN method.

Both classifiers achieve their best results on mammograms by selecting 10 components as features. But SVM has a higher recognition rate (87%) using polynomial kernel than KNN which has (79%) when the value of K is set to 5.

Mammogram patches dataset still increase the system performance. This is clearly obvious in Figure (4-21) and (4-22) that shows the results of SVM and KNN respectively.

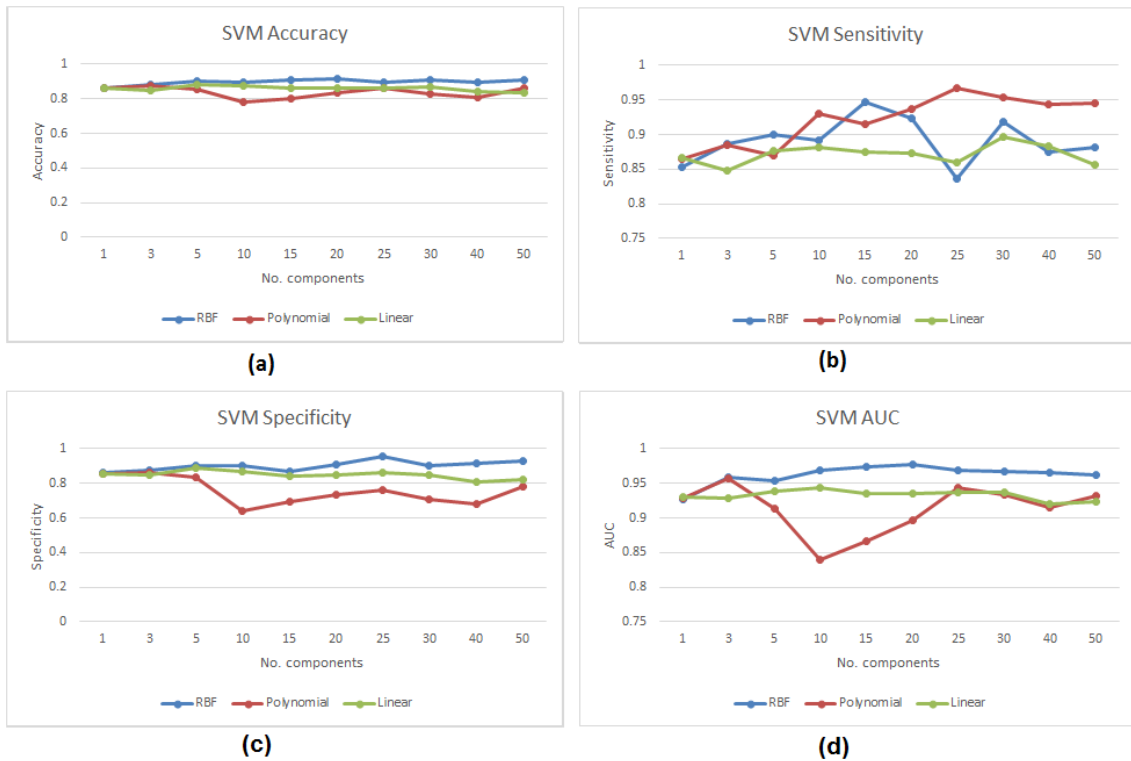


Figure 4-21: NMF results using mammogram patches and the SVM method.

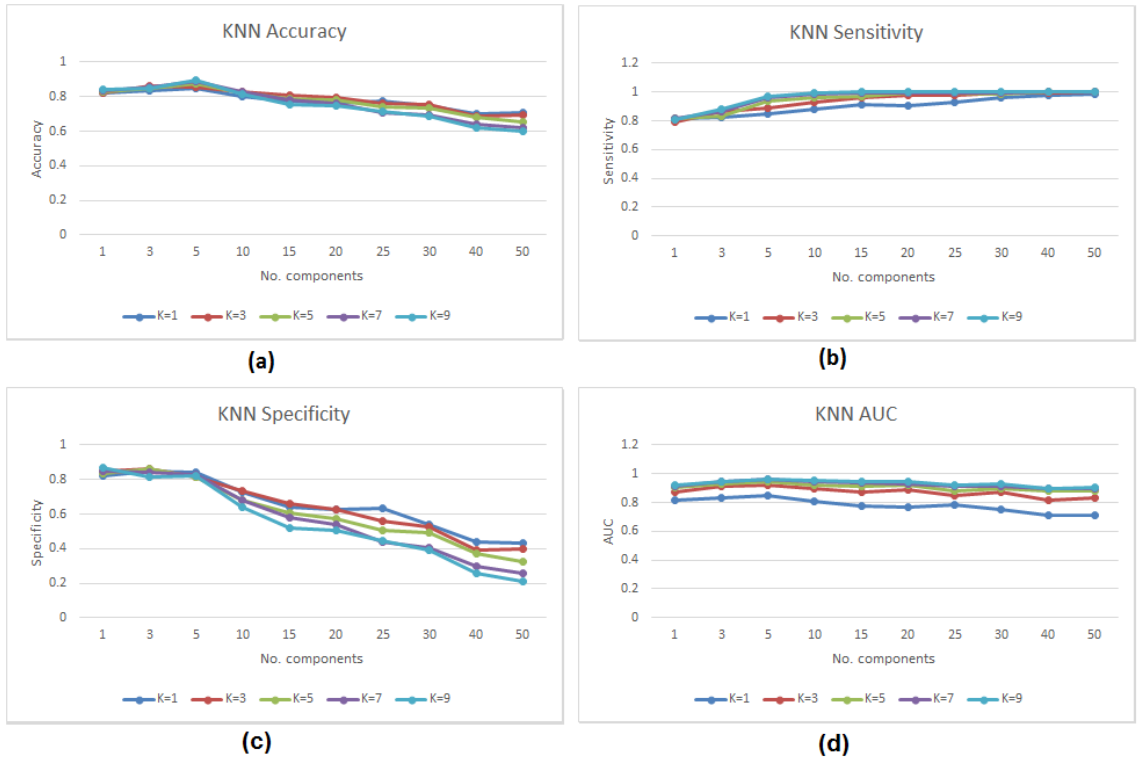


Figure 4-22: NMF results using mammogram patches and the KNN method.

The SVM performance achieves (92%) recognition rate with RBF kernel and 20 basis selected, whereas KNN achieves (90%) recognition rate with K set to 9 and 5 basis used with an increase of 11%.

The NMF feature extraction time is presented in Table (4-4)

Table 4-4: Estimated time for feature extraction using the NMF factorization.

	Mammograms	Patches
Feature Extraction Time (Seconds)	7.73	7.5

4.2.5 Mean Density Results

In this feature, the grey level threshold values used are 42, 180. Applying these threshold values results in segmenting the breast into three regions: the black area represents the background, the grey area represents fat and white area represents tissue. Figure (4-23) and (4-24) shows two samples of fatty and dense after segmentation.



Figure 4-23: A segmented fatty mammogram.



Figure 4-24: A segmented dense mammogram.

In the experiment, SVM with RBF kernel has the highest recognition rate (93%). The recognition results are presented in Table (4-5).

Table 4-5: SVM performance measures using full mammograms and the Mean Density method.

Kernel	Accuracy	Sensitivity	Specificity	AUC
RBF	0.93	0.87	0.98	0.93
Polynomial	0.92	0.88	0.96	0.95
Linear	0.84	0.94	0.75	0.95

The two classifiers have the same accuracy and they are close to each other in performance measures. Table (4-6) shows the obtained results from KNN.

Table 4-6: KNN performance measures using full mammograms and the Mean Density method.

KNN	Accuracy	Sensitivity	Specificity	AUC
1	0.87	0.86	0.88	0.87
3	0.89	0.87	0.91	0.92
5	0.91	0.88	0.95	0.93
7	0.92	0.88	0.96	0.94
9	0.92	0.88	0.96	0.94

This features does not affected by the type of database. Nevertheless, using mammogram patches dataset the results are close to the results of mammogram dataset as indicated by Table (4-7).

Table 4-7: SVM performance measures using mammogram patches and the Mean Density method.

Kernel	Accuracy	Sensitivity	Specificity	AUC
RBF	0.93	0.90	0.95	0.93
Polynomial	0.93	0.90	0.95	0.95
Linear	0.89	0.91	0.86	0.96

These results does not differ greatly from the results of mammograms. The two classifiers perform equally and differ slightly. The KNN obtained results are presented in Table (4-8).

Table 4-8: KNN performance measures using mammogram patches and the Mean Density method.

KNN	Accuracy	Sensitivity	Specificity	AUC
1	0.86	0.85	0.86	0.86

3	0.90	0.89	0.92	0.93
5	0.92	0.90	0.95	0.93
7	0.93	0.90	0.95	0.93
9	0.93	0.90	0.95	0.93

In addition, the feature extraction time for this feature is presented in Table (4-9).

Table 4-9: Estimated time for feature extraction using the Mean Density method.

	Mammograms	Patches
Feature Extraction Time (Seconds)	0.15	0.08

4.2.6 LBP Results:

This feature is one of the texture analysis techniques. Since it provide good results when combined with other features, in this work, this feature has been combined with NMF feature.

For each image in the dataset the NMF image is obtained and LBP features are extracted from NMF estimated image. Since NMF provide better results on the 10 components on mammograms dataset and 20 components on patches, these components are selected.

The SVM classifier achieves best results using linear kernel, the recognition rate dropped down from (87%) to (80%) as shown in Table (4-10).

Table 4-10: SVM performance measures using full mammograms and the NMF+LBP method.

Kernel	Accuracy	Sensitivity	Specificity	AUC

RBF	0.78	0.77	0.79	0.87
Polynomial	0.63	0.64	0.63	0.68
Linear	0.80	0.79	0.81	0.88

Similar to SVM, KNN recognition rate went down by 13% as shown in Table (4-11).

Table 4-11: KNN performance measures using full mammograms and the NMF+LBP method.

KNN	Accuracy	Sensitivity	Specificity	AUC
1	0.66	0.62	0.69	0.65
3	0.66	0.65	0.68	0.71
5	0.66	0.67	0.65	0.71
7	0.65	0.66	0.63	0.72
9	0.66	0.68	0.64	0.72

On the contrary to mammograms dataset, the performance increased in patches dataset using SVM. However, the NMF+LBP results does not reach the NMF results. Table (4-12) shows SVM performance measures. Undoubtedly the RBF kernel does a better job by achieving (92%) of accuracy.

Table 4-12: SVM performance measures using mammogram patches and the MF+LBP method.

Kernel	Accuracy	Sensitivity	Specificity	AUC
RBF	0.92	0.91	0.93	0.96
Polynomial	0.78	0.82	0.74	0.83
Linear	0.89	0.88	0.90	0.95

In addition, KNN achieves (83%) accuracy as shown in Table (4-13).

Table 4-13: KNN performance measures using mammogram patches and the MF+LBP method.

KNN	Accuracy	Sensitivity	Specificity	AUC
1	0.80	0.81	0.80	0.80
3	0.83	0.84	0.82	0.88
5	0.82	0.85	0.80	0.89
7	0.83	0.83	0.83	0.90
9	0.83	0.84	0.82	0.90

The feature extraction time is shown in Table (4-14).

Table 4-14: Estimated time for feature extraction using the NMF+LBP method.

	Mammograms	Patches
Feature Extraction Time (Seconds)	117	111.4

4.3 Conclusion

In the conclusion, the proposed system for breast density classification does what it was intended to do. The maximum average recognition rate obtained was 93%. The system achieves also good results compared to other systems in the literature as shown in Table (4-15).

Table 4-15: Benchmarking of the proposed approach.

	Accuracy of binary classification	Database used
2014 [38]	96%	MIAS
The proposed approach	93%	MIAS
2007 [30]	80%	MIAS
2009 [28]	90%	IRMA
2010 [25]	89%	MIAS + Trueta
2010 [26]	80%	IRMA
2010 [31]	80%	MIAS
2011 [40]	86%	MIAS
2011 [42]	90%	Oxford

We notice also that using patches increase the performance of the system in all cases except the Mean Density feature which does not affected by the database type.

Comparing all features together, the results are close to each other. The Mean Density features have the highest recognition rate. Unlike the other features, this features does not belong to the matrix factorization techniques. The simplicity of this feature makes it the fastest feature in case of computations. It does not need bigger memory or huge computations. However, this approach can be affected by many challenges, i.e. light. If the mammogram is bright, it can be misclassified. This can be shown clearly in some misclassified cases. Figure (4-25) presents a bright fatty mammogram after segmentation.

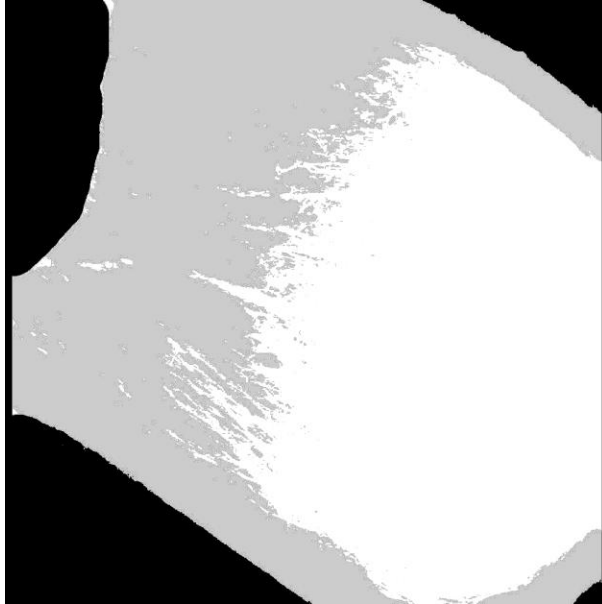


Figure 4-25: Misclassified brighter fatty mammogram.

Another dense mammogram with low bright is shown in Figure (4-26).

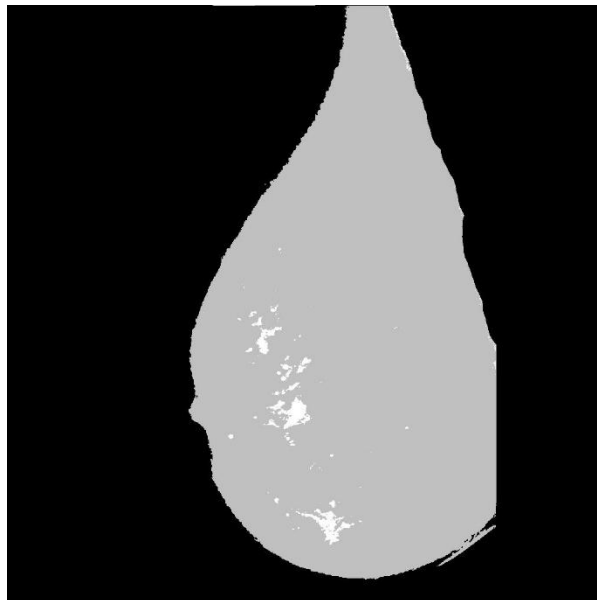


Figure 4-26: Misclassified darker dense mammogram.

In matrix factorization techniques, SVD has the ability to capture rich of information in a few basis. The 2D PCA also uses a small number of basis to capture most of the information. NMF technique uses more basis in information capturing, this is due to the non-negativity constrain applied when finding the factors using iterative approach. Table

(4-16) illustrates the factorization methods and the number of basis used to capture the most significant information in the images.

Table 4-16: Number of basis used in each method.

	PCA	2D PCA	SVD	NMF
Mammograms	47	11	5	100
Patches	34	6	1	50

From the experiments, the recognition rate of SVM classifier is better than KNN in most cases, due to the ability of customizing and changing of the kernel. Moreover, the RBF kernel has the ability to find a good separation that provide high recognition rate in most cases.

CHAPTER 5

CONCLUSION AND FUTURE WORK

In this section a summary of the contributions in this thesis is conducted and the limitations are highlighted to determine some directions for the future work.

A breast density classification technique is developed to address the problem of classifying breast densities according to BI-RADS lexicon. This technique achieves good accuracy considering two categories of densities – fatty and dense.

In this work, a literature review was presented in chapter 2 that shows different features and classification techniques used for breast density classification.

The design of the proposed system was discussed in chapter 3, the process starts by segmenting the mammograms and extract ROI (patches). Then a set of features are extracted from the mammograms and from the patches to show which improve the system performance. In this chapter six different features used are explained. Also the system uses two different classification approaches, each uses its own mechanism in the classification phase. These classifiers are reviewed also in this chapter.

We have conducted an intensive set of experiments on both mammograms and patches. We also compared the system performance of each feature using a set of performance measures. The experimental results are presented in chapter 4.

To our knowledge, the main contribution of this thesis is introducing the NMF and threshold as a new features that characterize the breast density, and the second

contribution is the combination between all features with different classifiers: SVM and KNN classifiers.

Although the promising results, the proposed system suffer from some drawbacks when using some features. PCA and SVD feature, requires more computation time and memory when obtaining the covariance matrix and finding the eigenvectors, while threshold feature is sensitive to the light in mammograms. The NMF is fast, however, it needs more basis to capture more information. The promising technique is the Mean Density. This technique requires less memory and less computations compared to other techniques.

5.1 Future Directions

These are some future trends and research directions in the breast density classification:

1. Construct a general database for research that provide more information.
2. Improve the system by considering the 4 categories of BI-RADS system.
3. Advance the system to the second step of discovering the breast cancer and tumor.
4. Using fusion and multi-classifiers may improve the system performance.

References

- [1] “American Cancer Society | Information and Resources for Cancer: Breast, Colon, Lung, Prostate, Skin.” [Online]. Available: <http://www.cancer.org/>. [Accessed: 25-Oct-2014].
- [2] “Comprehensive Cancer Information - National Cancer Institute.” [Online]. Available: <http://www.cancer.gov/>. [Accessed: 25-Oct-2014].
- [3] “Breastcancer.org - Breast Cancer Information and Awareness.” [Online]. Available: <http://www.breastcancer.org/>. [Accessed: 25-Oct-2014].
- [4] “Kingdom of Saudi Arabia - Ministry of Health Portal.” [Online]. Available: <http://www.moh.gov.sa/en/>. [Accessed: 25-Oct-2014].
- [5] “Zahra Breast Cancer Association.” [Online]. Available: <http://www.zahra.org.sa/Pages/default.aspx>. [Accessed: 25-Oct-2014].
- [6] “Breast Anatomy - National Breast Cancer Foundation.” [Online]. Available: <http://www.nationalbreastcancer.org/breast-anatomy>. [Accessed: 25-Oct-2014].
- [7] J. N. Wolfe, “Breast patterns as an index of risk for developing breast cancer,” *AJR Am. J. Roentgenol.*, vol. 126, no. 6, pp. 1130–1137, Jun. 1976.
- [8] V. Assi, J. Warwick, J. Cuzick, and S. W. Duffy, “Clinical and epidemiological issues in mammographic density,” *Nat. Rev. Clin. Oncol.*, vol. 9, no. 1, pp. 33–40, Jan. 2012.
- [9] V. A. McCormack and I. dos S. Silva, “Breast Density and Parenchymal Patterns as Markers of Breast Cancer Risk: A Meta-analysis,” *Cancer Epidemiol. Biomarkers Prev.*, vol. 15, no. 6, pp. 1159–1169, Jun. 2006.
- [10] R. W. Pinsky and M. A. Helvie, “Mammographic Breast Density: Effect on Imaging and Breast Cancer Risk,” *J. Natl. Compr. Canc. Netw.*, vol. 8, no. 10, pp. 1157–1165, Oct. 2010.
- [11] D. B. Kopans, *Breast Imaging*. Lippincott-Raven, 1998.
- [12] “WHO | Cancer.” [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs297/en/>. [Accessed: 25-Oct-2014].
- [13] J. Bozek, M. Mustra, K. Delac, and M. Grgic, “A Survey of Image Processing Algorithms in Digital Mammography,” in *Recent Advances in Multimedia Signal Processing and Communications*, M. Grgic, K. Delac, and M. Ghanbari, Eds. Springer Berlin Heidelberg, 2009, pp. 631–657.
- [14] J. S. Suri and R. M. Rangayyan, Eds., *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*. 1000 20th Street, Bellingham, WA 98227-0010 USA: SPIE, 2006.
- [15] R. Bronstein, D. Reyes and T. Kaufman, *Breast Imaging: Clinician's Guide to Diagnostic Imaging*, Springer New York, 2014.
- [16] W. A. Berg, Z. Zhang, D. Lehrer, R. A. Jong, E. D. Pisano, R. G. Barr, M. Böhm-Vélez, M. C. Mahoney, W. Phil Evans, L. H. Larsen, M. J. Morton, E. B. Mendelson, D. M. Farria, J. B. Cormack, H. S. Marques, A. Adams, N. M. Yeh, G. Gabrielli, and for the ACRIN 6666 Investigators, “Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography

- in women with elevated breast cancer risk,” *JAMA*, vol. 307, no. 13, pp. 1394–1404, Apr. 2012.
- [17] L. Liberman and J. H. Menell, “Breast imaging reporting and data system (BI-RADS),” *Radiol. Clin. North Am.*, vol. 40, no. 3, pp. 409–430, 2002.
- [18] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai, and H. N. Du, “Approaches for automated detection and classification of masses in mammograms,” *Pattern Recognit.*, vol. 39, no. 4, pp. 646–668, Apr. 2006.
- [19] M. J. Yaffe, “Mammographic density. Measurement of mammographic density,” *Breast Cancer Res. BCR*, vol. 10, no. 3, p. 209, 2008.
- [20] J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe, “The quantitative analysis of mammographic densities,” *Phys. Med. Biol.*, vol. 39, no. 10, p. 1629, Oct. 1994.
- [21] R. Highnam, S. M. Brady, M. J. Yaffe, N. Karssemeijer, and J. Harvey, “Robust Breast Composition Measurement - VolparaTM,” in *Digital Mammography*, J. Martí, A. Oliver, J. Freixenet, and R. Martí, Eds. Springer Berlin Heidelberg, 2010, pp. 342–349.
- [22] M. Jeffreys, J. Harvey, and R. Highnam, “Comparing a New Volumetric Breast Density Method (VolparaTM) to Cumulus,” in *Digital Mammography*, J. Martí, A. Oliver, J. Freixenet, and R. Martí, Eds. Springer Berlin Heidelberg, 2010, pp. 408–413.
- [23] C. Nickson, Y. Arzhaeva, Z. Aitken, T. Elgindy, M. Buckley, M. Li, D. R. English, and A. M. Kavanagh, “AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes,” *Breast Cancer Res.*, vol. 15, no. 5, p. R80, Sep. 2013.
- [24] Z. Chen, A. Oliver, E. Denton, and R. Zwiggelaar, “Automated Mammographic Risk Classification Based on Breast Density Estimation,” in *Pattern Recognition and Image Analysis*, J. M. Sanches, L. Micó, and J. S. Cardoso, Eds. Springer Berlin Heidelberg, 2013, pp. 237–244.
- [25] A. Oliver, X. Lladó, E. Pérez, J. Pont, E. R. E. Denton, J. Freixenet, and J. Martí, “A Statistical Approach for Breast Density Segmentation,” *J. Digit. Imaging*, vol. 23, no. 5, pp. 527–537, Oct. 2010.
- [26] J. E. de Oliveira, A. de A. Araújo, and T. M. Deserno, “MammoSysLesion: a content-based image retrieval system for mammographies,” in *17th International Conference on Systems Signals and Image Processing*, 2010.
- [27] T. M. Deserno, M. Soiron, J. E. E. de Oliveira, and A. De A Araujo, “Towards Computer-Aided Diagnostics of Screening Mammography Using Content-Based Image Retrieval,” in *2011 24th SIBGRAPI Conference on Graphics, Patterns and Images (Sibgrapi)*, 2011, pp. 211–219.
- [28] J. de Oliveira, A. Lopes, G. Camara-Chavez, A. de A. Araujo, and T. M. Deserno, “MammoSVD: A content-based image retrieval system using a reference database of mammographies,” in *22nd IEEE International Symposium on Computer-Based Medical Systems, 2009. CBMS 2009*, 2009, pp. 1–4.
- [29] J. E. E. de Oliveira, A. de Albuquerque Araujo, and T. M. Deserno, “Content-based image retrieval applied to BI-RADS tissue classification in screening mammography,” *World J. Radiol.*, vol. 3, no. 1, pp. 24–31, Jan. 2011.

- [30] H. S. Sheshadri and A. Kandaswamy, "Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms," *Comput. Med. Imaging Graph.*, vol. 31, no. 1, pp. 46–48, Jan. 2007.
- [31] L. Liu, J. Wang, and K. He, "Breast density classification using histogram moments of multiple resolution mammograms," in *2010 3rd International Conference on Biomedical Engineering and Informatics (BMEI)*, 2010, vol. 1, pp. 146–149.
- [32] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.
- [33] A. Oliver, J. Freixenet, and R. Zwigglelaar, "Automatic classification of breast density," in *IEEE International Conference on Image Processing, 2005. ICIP 2005*, 2005, vol. 2, pp. II–1258–61.
- [34] A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Perez, E. R. E. Denton, and R. Zwigglelaar, "A Novel Breast Tissue Density Classification Methodology," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 1, pp. 55–65, Jan. 2008.
- [35] T. Matsubara, D. Yamazaki, M. Kato, T. Hara, H. Fujita, T. Iwase, and T. Endo, "An automated classification scheme for mammograms based on amount and distribution of fibroglandular breast tissue density," *Int. Congr. Ser.*, vol. 1230, pp. 545–552, Jun. 2001.
- [36] T. S. Subashini, V. Ramalingam, and S. Palanivel, "Automated Assessment of Breast Tissue Density in Digital Mammograms," *Comput Vis Image Underst.*, vol. 114, no. 1, pp. 33–43, Jan. 2010.
- [37] S. Petroudi, I. Constantinou, C. Tziakouri, M. Pattichis, and C. Pattichis, "Investigation of AM-FM methods for mammographic breast density classification," in *2013 IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2013, pp. 1–4.
- [38] V. Sharma and S. Singh, "CFS–SMO based classification of breast density using multiple texture models," *Med. Biol. Eng. Comput.*, vol. 52, no. 6, pp. 521–529, Jun. 2014.
- [39] N. Vález, G. Bueno, O. Déniz, J. Dorado, J. A. Seoane, A. Pazos, and C. Pastor, "Breast density classification to reduce false positives in CADe systems," *Comput. Methods Programs Biomed.*, vol. 113, no. 2, pp. 569–584, Feb. 2014.
- [40] Z. Chen, E. Denton, and R. Zwigglelaar, "Local feature based mammographic tissue pattern modelling and breast density classification," in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, 2011, vol. 1, pp. 351–355.
- [41] G. Liasis, C. Pattichis, and S. Petroudi, "Combination of different texture features for mammographic breast density classification," in *2012 IEEE 12th International Conference on Bioinformatics Bioengineering (BIBE)*, 2012, pp. 732–737.
- [42] S. Petroudi and M. Brady, "Breast density characterization using texton distributions," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, 2011, pp. 5004–5007.
- [43] *Foundations of Image Understanding*.
- [44] Y. Javed and M. M. Khan, "Image texture classification using textons," in *2011 7th International Conference on Emerging Technologies (ICET)*, 2011, pp. 1–5.

- [45] Y. Javed, M. M. Khan, and J. Chaussoot, "Population density estimation using textons," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, 2012, pp. 2206–2209.
- [46] G. N. Srinivasan and G. Shobha, "Statistical texture analysis," in *Proceedings of world academy of science, engineering and technology*, 2008, vol. 36, pp. 1264–1269.
- [47] T. Leung and J. Malik, "Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 29–44, Jun. 2001.
- [48] J. Zhang, H. Zhao, and J. Liang, "Continuous Rotation Invariant Local Descriptors for Texton Dictionary-based Texture Classification," *Comput Vis Image Underst*, vol. 117, no. 1, pp. 56–75, Jan. 2013.
- [49] "Visual Geometry Group - Texture Classification." [Online]. Available: <http://www.robots.ox.ac.uk/~vgg/research/texclass/with.html>. [Accessed: 26-Oct-2014].
- [50] A. Bosch, X. Muoz, A. Oliver, and J. Marti, "Modeling and Classifying Breast Tissue Density in Mammograms," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, vol. 2, pp. 1552–1558.
- [51] S. D. Tzikopoulos, M. E. Mavroforakis, H. V. Georgiou, N. Dimitropoulos, and S. Theodoridis, "A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry," *Comput. Methods Programs Biomed.*, vol. 102, no. 1, pp. 47–63, Apr. 2011.
- [52] Q. Liu, L. Liu, Y. Tan, J. Wang, X. Ma, and H. Ni, "Mammogram density estimation using sub-region classification," in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, 2011, vol. 1, pp. 356–359.
- [53] W. R. Silva and D. Menotti, "Classification of Mammograms by the Breast Composition," *Proc Image Process. Comput. Vis.*, pp. 58–63, 2012.
- [54] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.
- [55] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," in *Computational Statistics and Data Analysis*, 2006, pp. 155–173.
- [56] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.
- [57] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [58] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis," *Neural Comput.*, vol. 21, no. 3, pp. 793–830, Sep. 2008.
- [59] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1055–1064, Sep. 1999.
- [60] J. A. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, J. A. K. Suykens, and T. Van Gestel, *Least squares support vector machines*, vol. 4. World Scientific, 2002.

- [61] “NMF:DTU Toolbox.” [Online]. Available:
<http://cogsys.imm.dtu.dk/toolbox/nmf/>. [Accessed: 27-Oct-2014].
- [62] F. Gorunescu, “Classification Performance Evaluation,” in *Data Mining*, Springer Berlin Heidelberg, 2011, pp. 319–330.

Vitae

Name : Abdullah Hussein Abdullah Owaidh |

Nationality : Yemeni |

Date of Birth : 4/15/1981 |

Email : ajop4@hotmail.com |

Address : Hadramout- Yemen |

Academic Background :

- Received Bachelor degree from Hadramout University of Science and Technology, Yemen, 2006.
- Joined King Fahd University of Petroleum and Minerals, October 2010 as Master student.
- Completed M.Sc. in Information and Computer Science from King Fahd University of Petroleum and Minerals, Dhahran, KSA, December 2014.

Publications :

- Ghouti, Lahouari, and Abdullah H. Owaidh. "NMF-Density: NMF-Based Breast Density Classifier." 22nd European Symposium on Artificial Neural Networks (ESANN 2014), pp.455-460, 23-25 April 2014, Bruges, Belgium.

Patents :

- L. Ghouti and A. H. Owaidh, "Method, System and Computer Program product for Breast Density Classification Using Part-Based Local Features," US Patent Application, Docket No. ICS899, November 2014.