

**ROBUST TUMOR DETECTION IN MAMMOGRAM
IMAGES USING AN OPTIMAL SET OF TEXTURE
FEATURES**

BY

MAHMOUD M. M. ALHELOU

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

1963 ١٣٨٣

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

ELECTRICAL ENGINEERING

June, 2016

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

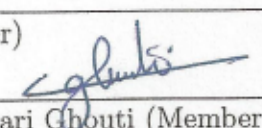
DEANSHIP OF GRADUATE STUDIES

This thesis, written by MAHMOUD M. M. ALHELOU under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

Thesis Committee



Dr. Mohamed Elariche (Adviser)

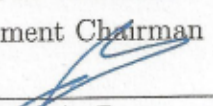
(Co-adviser)


Dr. Lahouari Ghouti (Member)

Dr. Mohamed Mohandes (Member)


(Member)


Dr. Ali A. Al-Shaikhi
Department Chairman


Dr. Salam A. Zummo
Dean of Graduate Studies

Date

22/6/16



©Mahmoud M. M. Alhelou
2016

Dedications

*To the soul of my father,
and my beloved mother.*

ACKNOWLEDGMENTS

First of all, my deepest and sincere gratitude to almighty Allah for his infinite and persistent help and blessings. I would like to express my gratefulness to my late father who has always inspired me to become the man who I am. I wish to thank my beloved mother for her unconditional love and all of my brothers and sisters for standing always beside me.

Many thanks to my adviser and academic godfather Dr. Mohamed Deriche for his appreciated support and valuable guidance that helped me finish this work. My profound gratitude goes to my thesis committee members Dr. Lahouari Gouti and Dr. Mohamed Mohandes for providing me with their valuable comments.

My genuine thanks to King Faisal Specialist Hospital & Research Centre staff specially Dr. Nuha Khumais for her worthy consultancy that helped me in understanding the problem of breast cancer detection from the perspective of medical professionals. I would like to thank Dr. Ahmed Shamia and Dr. Ahmed Alashqar for their precious feedback and meetings that helped me in developing my knowledge and insight of the breast cancer detection challenges.

Finally, I acknowledge my colleges and wonderful friends Mr. Mohammad Qureshi and Mr. Asjad Amin for supporting and encouraging me through out all of this work.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF ABBREVIATIONS	ix
ABSTRACT (ENGLISH)	xii
CHAPTER 1 INTRODUCTION	1
1.1 Background	1
1.2 Problem statement	3
1.2.1 Research objectives	4
1.3 Main contributions	5
1.4 Organization of the thesis	6
CHAPTER 2 LITERATURE REVIEW	7
2.1 CAD approaches for breast density classification	8
2.2 CAD approaches for cancer detection	13
2.3 Texture analysis techniques	21
2.3.1 Local Binary Patterns (LBP) texture features	22
2.3.2 Gray Level Co-occurrence Matrix (GLCM) features	24
2.3.3 Gray Level Run Length Matrix (GLRL) features	30
2.3.4 Wavelet based texture analysis	33
2.3.5 Gabor texture analysis	36

2.3.6	Summary	37
CHAPTER 3 RESEARCH METHODOLOGY		38
3.1	Proposed approach	38
3.1.1	Image Preprocessing	40
3.1.2	Feature extraction	41
3.1.3	Feature selection	43
3.1.4	Pattern Classification	44
3.2	System Implementation	48
3.2.1	Image Databases	48
3.2.2	Experimental setup	50
CHAPTER 4 RESULTS AND DISCUSSION		54
4.1	Density Classification	54
4.2	Cancer detection	59
4.3	The proposed two stage classifier	62
CHAPTER 5 CONCLUSION AND FUTURE WORK		65
5.1	Conclusion	65
5.2	Future work	66
REFERENCES		67
VITAE		80

LIST OF TABLES

2.1	BIRADS Classification	8
2.2	Brief summary of work on the DDSM database	9
2.3	Brief summary of work on the MIAS database	10
2.4	Summary of popular feature extraction and classification methods . . .	14
2.5	Basic notation used in GLCM computations	26
4.1	Density classification using all features	56
4.2	Density classification using feature selection	57
4.3	The top 20 features for 4 BIRADS classes	57
4.4	The top 20 features for low and high densities	58
4.5	Comparison for density classification with four BIRADS classes	58
4.6	Comparison for density classification with low and high densities classes	59
4.7	Classification rate for cancer detection for all images (BIRADS I,II,III,IV) (without density information , all features)	60
4.8	Classification rate for Cancer detection based on Low and High densities for setup 1 (Normal ,benign and malignant)(known densities, selected features).	60
4.9	Classification rate for Cancer detection based on Low and High densities for setup 2 (Cancer/No cancer)(known densities,selected features). . . .	61
4.10	The optimal features for cancer classification using High and Low den- sity classes	61
4.11	Summery of 2-stage classifier	64

LIST OF FIGURES

1.1	A typical Mammography Device [1]	3
1.2	An example of showing the two views(MLO and CC) with different regions of intrest [2]	4
2.1	Overall CAD systems used in mammogram analysis	7
2.2	Examples of BIRADS examples	8
2.3	A full view of the approach proposed by [3].	12
2.4	Overall structure of CAD cancer detection systems	15
2.5	An example of GLCM constructed for a small patch from a mammo-gram image [4].	16
2.6	The central pixel g_c and its P circularly symmetric neighbors at radius R	17
2.7	Multiscale texture modeling [5] : (a) image marked by a radiologist, (b)Detect abnormal ROIs, (c) Classify ROIs into AD/non-AD, and (d) Detected correct ROI based on (c).	20
2.8	LBP for Circular neighborhood [6]	22
2.9	An example of finding LBP with R=1,P=8 and gray level value $g_c =70$	24
2.10	An example of calculating GLCM for 0°	25
2.11	The structure of a 2D Wavelet decomposition with 1 Level [7]	34
2.12	The structure of a 2D Wavelet decomposition with 2 Levels [8]	35
2.13	An example for 2D Wavelets decomposition with 2 Levels	35
2.14	An example of Gabor filter masks over five scales and eight orientations	37
3.1	Full view of the proposed CAD system	39
3.2	Preprocessing steps	40

3.3	An example of applying a sequence of preprocessing steps [9]	40
3.4	An example of applying a sequence of preprocessing steps: (a) the original image, (b) after applying label removal, and (c) after applying pectoral removal	41
3.5	Density classification	47
3.6	Cancer classification	47
3.7	MIAS examples images	49
3.8	IRMA example images	50
3.9	Preprocessing example	51
4.1	IRMA example images	55
4.2	Experiments developed for our proposed algorithm	63

LIST OF ABBREVIATIONS

Abbreviation	Description
ASR	Age-Standardized Rate
ACR	American College of Radiology
MLO	Mediolateral Oblique View
CC	Craniocandal View
CAD	Computer-Aided Diagnosis
ROI	Region Of Interest
BIRADS	Breast Imaging-Reporting and Data System
MIAS	The Mammographic Image Analysis Society
DDSM	The Digital Database for Screening Mammography
MR8	Maximum Response 8
GLCM	Gray Level Co-occurrence matrix
kNN	K-Nearest Neighbors classifier
SIFT	Scale Invariant Feature Transforms
LVQ	Learning Vector Quantization
SVM	Support Vector Machine
LBP	Local Binary Patterns
LGA	Local Greylevel Appearances
BIF	Basic Image Features
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis

NMF	Non-negative Matrix Factorization
GLRL	Gray Level Run Length Matrix
GLLD	Gray Level and Local Difference
ANN	Artificial Neural Networks
FS	Forward Selection
BS	Backward Selection
GMM	Gaussian Mixture Model
GT	Ground Truth
ROC	Receiver Operating Characteristic
AUC	The Area Under the (ROC) Curve
SRE	Short Run Emphasis
LRE	Long Run Emphasis
GLN	Gray-Level Non-uniformity
RLN	Run Length Non-uniformity
RP	Run Percentage
LGRE	Low Gray-Level Run Emphasis
HGRE	High Gray-Level Run Emphasis
SRLGE	Short Run Low Gray-Level Emphasis
SRHGE	Short Run High Gray-Level Emphasis
LRLGE	Long Run Low Gray-Level Emphasis
LRHGE	Long Run High Gray-Level Emphasis
IRMA	Image Retrieval in Medical Applications

IG	Information Gain
SMO	Sequential Minimal Optimization
LLNL	Lawrence Livermore National Laboratories
CR	Classification Rate

THESIS ABSTRACT

NAME: Mahmoud M. M. Alhelou

TITLE OF STUDY: Robust tumor detection in mammogram images using an optimal set of texture features

MAJOR FIELD: Electrical Engineering Department

DATE OF DEGREE: June 22, 2016

Breast cancer is one of the most common types of cancer, as well as the leading cause of mortality among women. Recent statistics showed that around 25% of cancer cases among women are of breast type. Mammography is currently the most effective imaging modality for the detection of breast cancer and diagnosis of abnormalities which can identify cancerous cells. The problem of early detection of cancer from mammographic images is still a major challenge and performance is still not acceptable by medical professionals. Here, we propose to develop a new image processing approach for early detection of cancer from mammogram images using robust texture features combined with a statistical classifier. The aim is to start with a pool of features then use a feature selection technique to extract the most relevant features for the detection of cancerous image patches. A new two stage classifier is developed which combines

density classification with cancer classification. The first stage of the proposed algorithm achieved a density classification of 93.56% from two BIRADs. Cancer/no cancer classification is carried during the second stage. The hybrid two stage classifier achieved an overall classification rate of 80%. The results were validated on both the MIAS and IRMA databases. We also show that the reduced size optimal set of features using small patches achieves excellent results in both density classification and cancer detection. A comparison to state-of-the-art has been carried and showed that our approach provides an improvement both in terms of density classification and cancer detection accuracy.

ملخص الرسالة

الاسم الكامل: محمود مروان محمود الحلو

عنوان الرسالة: الإكتشاف الأمثل لسرطان الثدي من صور 'mammogram' باستخدام أفضل الخصائص المستخرجة من الصور

التخصص: الهندسة الكهربائية

تاريخ الدرجة العلمية: 20- يونيو - 2016

يعتبر سرطان الثدي من أكثر أنواع السرطان انتشارا عند النساء وهو أحد أهم أسباب الوفاة للمصابات بهذا المرض. أظهرت النتائج مؤخرا أن حوالي 25% من المرضى المصابات بالسرطان مصابات بسرطان الثدي، ويعتبر التصوير باستخدام صور Mammogram من أهم وأنجع الطرق للاكتشاف المبكر لسرطان الثدي في الوقت الحالي كما تستخدم لاكتشاف أي كتلة غير طبيعية موجودة في ثدي المرأة سواء كانت هذه الكتل عبارة عن خلايا سرطانية أو غيرها. في الوقت الحاضر تعتبر عملية الاكتشاف المبكر لسرطان الثدي عند النساء من أكثر وأصعب الاشكاليات التي تواجه الباحثين حيث أنهم لم يتوصلوا إلى الدقة المقبولة للمختصين في المجال الطبي. في هذه الرسالة قمنا بتطوير خوارزمية جديدة باستخدام معالجة الصور الرقمية واستخراج خصائص مميزة لكل صورة واستخدام أدوات التصنيف الاحصائية للحصول على نتائج جيدة. الهدف من هذه الخوارزمية استخراج الخصائص من الصور المستخدمة واختيار أفضل الخصائص التي تساعد على اكتشاف وجود الخلايا السرطانية من عدمه باستخدام أدوات التصنيف المناسبة. خلال هذا العمل تم تطوير خوارزمية مكونة من مرحلتين: في المرحلة الأولى يتم تصنيف الصورة حسب كثافة الأنسجة داخل الثدي حسب نظام BIRADS إلى نوعين: كثافة خفيفة و كثافة عالية وقد حصلنا على دقة تصنيف صور بنسبة 93.56%. في المرحلة الثانية يتم تصنيف الصورة حسب وجود الخلايا السرطانية من عدمه في كل صورة بعد انتهاء المرحلة الأولى وقد حصلنا على دقة كلية للمرحلتين بنسبة 80% و للحصول على هذه النتائج تم استخدام قاعدتي بيانات هما MIAS و IRMA. خلال هذه الرسالة تم استنتاج أنه حين يتم حذف بعض الخصائص التي لا تساعد المصنفات في عمليات التصنيف فإنه تزيد دقة التصنيف بشكل جيد. كما تم استخراج الخصائص التي تساعد على التصنيف بشكل جيد في كلتا المرحلتين. في ختام الرسالة تم عرض بعض الدراسات المشابهة لهذا العمل ومقارنة النتائج التي تحصلوا عليها مع ما توصلنا إليه من نتائج وتم ايضاح الفروقات.

CHAPTER 1

INTRODUCTION

1.1 Background

Breast cancer is the most common type of cancer among women over the world. The incidence of breast cancer has increased over the last two decades due to unhealthy lifestyles, stress, artificial foods ingredients, and other environmental factors. Therefore, the early detection is important to control the increase in breast cancer incidence [10].

In Saudi Arabia, a breast cancer ratio of 25.1%, among females, was reported for all newly diagnosed cases (5, 205) in year 2009. The five regions with the highest ASR (Age-Standardized Rate) were: the Eastern Region at 33.1/100,000, Riyadh Region at 29.4/100,000, Makkah Region at 26.4/100,000, Qassim Region at 25/100,000 and finally the Jouf Region at 22.5/100,000. The median age at diagnosis was 48 years [11].

In 2015, the American Cancer Society reported 231, 840 new cases of breast cancer

in the U.S. An estimated number of 40,290 women are expected to succumb from breast cancer [12][13].

To date, the most common diagnosis tool used at hospitals/clinics in detecting breast cancer is mammogram analysis. Mammogram images are simply X-ray images of the breast acquired by mammogram devices such as the one in Figure 1.1. For each mammogram image, we have two views: Mediolateral Oblique View (MLO) and Cranio-candal View (CC), and both views are used by the radiologist to detect the masses or any other abnormalities within the mammogram image. An example of a mammogram image is displayed in Figure 1.2 showing different regions of interest. Health professionals recommend that women over 40 years old should take the mammogram test once a year to avoid late detection of cancer. While acquiring mammographic images is not very difficult, their analysis on the other hand can be very challenging. The problem is that cancer textures can look very similar to surrounding breast tissue. The challenge for the medical professionals is to make a decision on whether a certain area/region is a cancer or just regular tissue. The breast density is seen as the main factor in any Computer aided diagnosis (CAD) system as high density breast requires additional processing like segmentation to differentiate between normal dense tissue and cancer masses. So, breast density classification is seen as a crucial preprocessing stage in cancer detection. To reduce the potential of misdiagnosis, patients may be subjected to unnecessary biopsy which is costly and stressful to patients. Computer aided diagnosis systems have been developed to close this gap in helping medical professionals in focusing only on suspicious regions of the mammogram images. However,

the detection of such suspicious regions itself is a challenge. This challenge prompted us to investigate the problem of developing a set of features that are optimal for the accurate classification of textures and hence detecting cancer in early stages. While there are hundred of features that can be extracted from mammographic images, many may be irrelevant to the task of cancer detection. Here, we develop a informative-theoretic framework to rank features and select an optimal limited size feature set. These features are then used not under a threshold framework but under a probabilistic framework to reach a final decision on whether a certain mammographic image contains cancerous regions or not.



Figure 1.1: A typical Mammography Device [1]

1.2 Problem statement

Breast cancer is a common cancer among females all over the world. The main challenge in mammogram analysis is the accurate detection of malignant tumors. More-

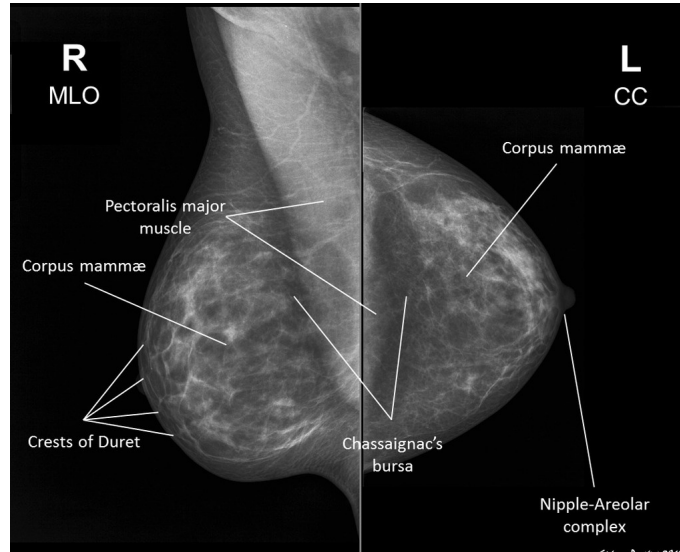


Figure 1.2: An example of showing the two views(MLO and CC) with different regions of interest [2]

over, the increased density in the breast area has added another challenge to cancer detection. As such, a robust mammogram analysis system needs to include both accurate identification of breast density followed by cancer detection.

Computer-Aided Diagnosis (CAD) system continues to be the main tools that can help radiologists in processing the substantial number of mammographic images while focusing only on suspicious regions. Here, we approach the problem of CAD for mammography using a 2-stage analysis system which first identifies the type of breast density, then uses a dedicated density-dependent classifier to identify the existence of cancerous regions.

1.2.1 Research objectives

The aim of this research is to introduce a new CAD system for breast density classification and cancer detection from mammograms using texture features. The work

includes the following components:

- Development of a preprocessing algorithm for pectoral removal using a region growing technique.
- Development of a method for breast density segmentation and Region Of Interest (ROI) detection using texture analysis.
- Development of a novel approach for breast density classification using the Breast Imaging-Reporting and Data System (BIRADS) databases.
- Development of a density-depending cancer detection algorithm using an optimal set of textures features.
- Extensive testing of the developed algorithms on public databases and comparison with existing techniques.

1.3 Main contributions

The main contributions of the thesis are:

- Development of a robust segmentation algorithm for pectoral region detection for mammogram images.
- Development of an optimal set of features using information theory from mammogram patches which is shown to improve breast density classification and cancer detection.

- Development a new framework for cancer detection using a 2-stage density/cancer classifier.
- Extensive experimental work on established mammographic databases showing the superior performance of the developed algorithms compared to existing approaches.

1.4 Organization of the thesis

In Chapter 2, we discuss the previous work carried in mammographic image analysis. In section 2.1, we discuss work related to density classification followed by work on cancer detection in section 2.2. In chapter 3, we present the methodology used for our work and discuss the details of the proposed algorithms. We then discuss the experimental results of our approach in chapter 4. Finally, we conclude our work and discuss future research directions in chapter 5.

CHAPTER 2

LITERATURE REVIEW

Substantial research efforts have been put in the development of algorithms for the detection of breast cancer from mammographic images. Common breast cancer detection stages consist of two main stages: Breast density classification followed by cancer classification as shown in Figure 2.1.

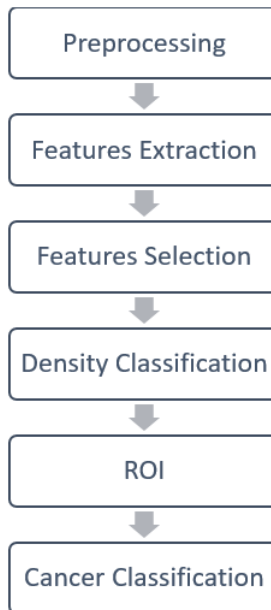


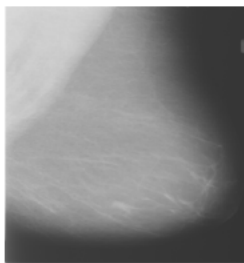
Figure 2.1: Overall CAD systems used in mammogram analysis

2.1 CAD approaches for breast density classification

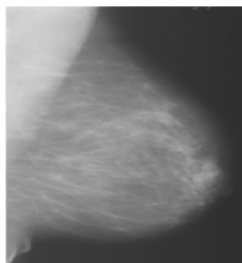
Recently, Birdwell et al [14] found that breast density was the second most common reason for missed cancer detection when evaluating causes of missed breast cancer. The American College of Radiology (ACR) [12] came up with standard that categorizes breast density into 7 so-called BIRADS. Most of the research is based on BIRAD I (fatty tissue) to BIRAD IV (extreme dense tissue). The details of these BIRADS are shown in Table 2.1 and example images are shown in Figure 2.2. The discussion on the seven classes can be found in [12].

Table 2.1: BIRADS Classification [15]

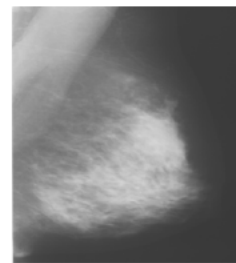
BI-RADS Class	Density(%)	Breast density
TYPE-I	00-25	Entirely fatty
TYPE-II	26-50	Some fibro-glandular tissue
TYPE-III	51-75	Heterogeneously dense breast
TYPE-IV	76-100	Extremely dense



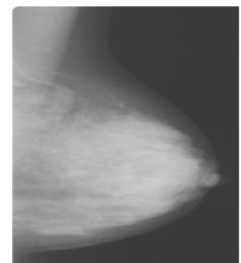
(a) BIRADS I



(b) BIRADS II



(c) BIRADS III



(d) BIRADS IV

Figure 2.2: Example of BIRADS images [15]

Kumar et al. [15] published a recent review paper related to different approaches developed for breast density classification. They found that two types of images are used for density classification : Segmented breast tissue and predefined ROI. For segmented breast tissue, the researchers used the full mammogram images and extracted a small patch to carry the processing over it. They discussed the developed approaches that used popular databases like the Mammographic Image Analysis Society (MIAS) [16] and the Digital Database for Screening Mammography (DDSM) [17]. Different approaches discussed in their review paper are shown in Table 2.2 for the DDSM and Table 2.3 for the MIAS databases.

Table 2.2: Brief study carried out on DDSM [15]

Author(s), Year	Segmented Breast Tissue/ROI	No. of im- ages	Classifier	Accuracy(%)	Considered Class
Oliver,et al. 2005	SBT	300 R-MLO	KNN, Deci- sion tree	47.00	BIRADS I- IV
Oliver,et al. 2008	SBT	831	SFS+KNN	77.00	BIRADS I- IV
Bovis,et al. 2002	SBT	377	ANN	71.40	BIRADS I- IV
Bosch,et al. 2006	SBT	500	KNN,SVM	84.75	BIRADS I- IV
Oliver,et al. 2010	SBT	831	LDA-PCA	79.00	Fatty,dense

Petroudi et al. [18] proposed a method for breast density classification based on the statistical distribution over textons. They worked on full mammogram images from the Oxford Database, segmented into patches, and features were extracted using Maximum Response 8 (MR8) filters. They classified patches into one of the four BIRADS classes with an overall classification rate of 75%, and when two classes only

Table 2.3: Brief summary of work on the MIAS database [15]

Author(s), Year	Segmented Breast Tis- sue/ROI	No. of images	Classifier	Accuracy(%)
Muhimmah,et al. 2006	SBT	321	DAG-SVM	77.57
Oliver,et al. 2005	SBT	270	KNN-Decision tree	67.00-73.00
Oliver,et al. 2008	SBT	43	SVM	95.55
Subashini,et al. 2010	SBT	377	ANN	71.40
Sharma,et al. 2014	ROI	322	SMO	96.46
Tzikopoulos,et al. 2010	SBT	322	SVM	85.70
Blot,et al. 2001	SBT	265	KNN	65.00
Qu,et al. 2011	SBT	322	E-FELM	72.67
Bosch,et al. 2006	SBT	322	KNN,SVM	95.42
Z. Chen,et al. 2011	SBT	322	KNN,Bayesian	75.00
Oliver,et al. 2010	SBT	322	LDA-PCA	94.00
Mustra,et al. 2012	ROI	322	KNN,Naive Bayesian	82.00

Note-SBT:Segmented breast tissue

low and high densities are used, an overall classification rate of 91% was achieved.

Oliver et al. [19] discussed different strategies for feature extraction based on the BIRADS classification. They used the full MIAS database and classified images into four classes BIRADS using a segmentation method for achieving an overall classification rate of 82%.

He et al. [20] developed an automatic mammographic density segmentation approach using a Bayes classifier. They used the MIAS database with full mammogram

images. They classified the images into one of the four BIRADS with a classification rate of 85% , and into low and high densities with a classification rate of 78%.

Mustra et al. [21] investigated different feature selection algorithms. They used the MIAS database with three categories: fatty, fatty-glandular, and dense-glandular. They obtained the Gray Level Co-occurrence matrix (GLCM) for the ROIs and used the K-Nearest Neighbors classifier (KNN). They achieved around 70% classification accuracy.

Oliver et al. [22] developed a new approach for breast parenchymal density to classify mammogram images using the four BIRADS classes. They identified breast density based on the combination of texture and gray level information. They started with extracting the breast area only from the full mammogram image using the DDSM database. They used the Fuzzy C-Means algorithm to segment the different tissue types, then classified these using KNN and decision trees into BIRADS classes. They achieved a 47% correct classification which is considerably low compared to other approaches.

Kutluk and Günsel [23] proposed an approach for classification of breast tissue density from mammographic images into one of 3 classes: fatty , fatty-glandular and dense-glandular. They used the Scale Invariant Feature Transform (SIFT) as a local feature extraction technique and the Learning Vector Quantization (LVQ) algorithm for supervised classification achieving a classification accuracy of over 90%.

Liasis et al. [3] investigated different texture features and their combination together with a Support Vector Machine (SVM) algorithm for breast density classi-

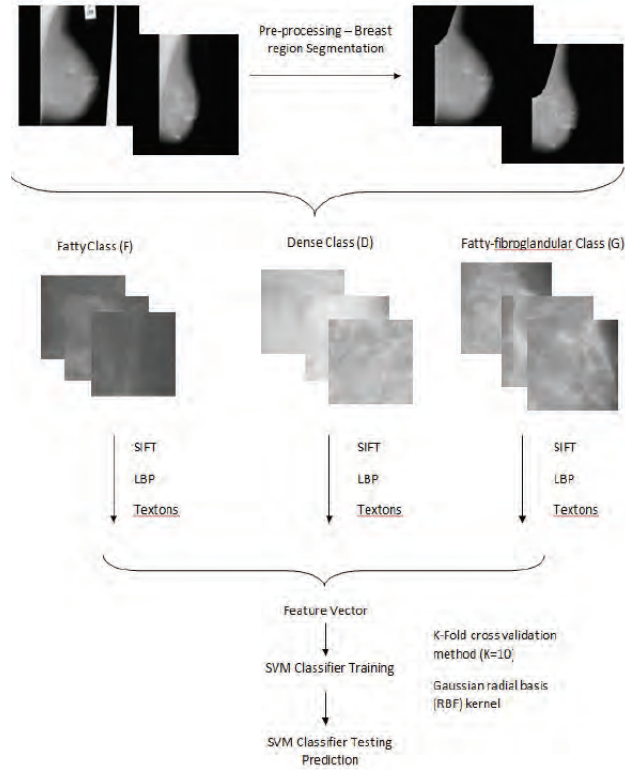


Figure 2.3: A full view of the approach proposed by [3].

fication as shown in Figure 2.3. SIFT, Local Binary Patterns (LBP), and texton histograms resulted in an excellent classification accuracy of 93.4% over the MIAS database.

Chen et al. [24] presented a model for mammographic tissue analysis in the form of feature histograms using five different types of local features: LBP, Local Greylevel Appearances (LGA), Textons I,II and Basic Image Features (BIF). They classified the images for the MIAS database into four BIRADS classes and into two low and high density classes. For four BIRADS classes, the overall classification accuracies achieved were 59%, 72%, 69%, 75% and 70% for LBP, LGA, Texton I, Texton II and BIF, respectively. For two classes, the accuracies were 79%, 86%, 83%, 88% and 86%, respectively.

Oliver et al. [25] presented an automatic evaluation approach for breast density considering fatty or dense classes. Their work was based on statistical analysis over patches extracted from mammogram images from the MIAS and the Trueta databases. They used Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) Models. They achieved a classification accuracy of 91%.

Ghouthi and Owaidh [26] developed new features to improve breast density classification. They extracted features from Non-negative Matrix Factorization (NMF) of patches of sizes 300×300 and 1024×1024 , and used an SVM classifier. They applied NMF on the MIAS database. Finally, PCA was applied on the extracted features achieving an accuracy of 83%.

As can be seen from the above analysis, density classification is considered to be a major step for improving cancer detection accuracy. Many researchers are simplifying this step by taking the two extreme situations: fatty and dense breast tissues. On the other hand, some researchers ignore this step by working on cancer detection for the same density class. Many approaches were discussed using different types of feature which may sometimes be irrelevant. This prompted us to develop a framework for finding the top ranked features relevant to the task of accurate density classification.

2.2 CAD approaches for cancer detection

In a recent work by Ganesan and his team, a comprehensive survey of different CAD approaches for cancer detection was presented. The authors explained that most existing CAD systems were not yet conclusive enough to warrant their credible clinical

usage. They also show, however, that the detection accuracy of cancer has improved with the introduction of advanced CAD-based diagnostic procedures. They tabulated different feature extraction and classification methods as shown in Table 2.4. From the Table 2.4, the best results are around 90%, however, these are not recommended yet for use in clinical trials [27].

Table 2.4: Summary of popular feature extraction and classification methods [27]

Author(s)	Method used	Accuracy(%)
Kimme,et al.	Normalized statistics and texture features	74
Petrosian,et al.	Spatial Gray Level Dependence and textural features with a decision tree classifier	76-89
Kinoshita,et al.	Shape and texture features with a three layer feed-forward neural network	81
Rangayyan,et al.	Region based edge-profile acutance measure	92
Polakowski,et al.	Model based vision algorithm. Difference of Gaussians and texture features	92
Priebe,et al.	Fractal texture measures	88
Sameti,et al.	Optical density, photometric and textural features	72
Chitre,et al.	Texture measures with artificial neural network	87
Mudigonda,et al.	Gray level co-occurrence matrices, polygonal modeling with jack-knife classification	83
Brijesh,et al.	Statistical features with fuzzy neural network	83
Yoshida,et al.	Wavelet features in combination with a difference image technique	90
Liyang Wei,et al.	Statistical features in a multiple view mammogram with SVM and KFD	85
Oliver A,et al.	Eigen faces approach	82-90
Szekeley,et al.	Texture features and a combining classifier of decision trees and multiresolution markov random models	88-94
Alolfe,et al.	Forward stepwise linear regression method with a combined classifier of SVM and LDA	82.5-90

For most existing approaches, the main stages for cancer detection are shown in

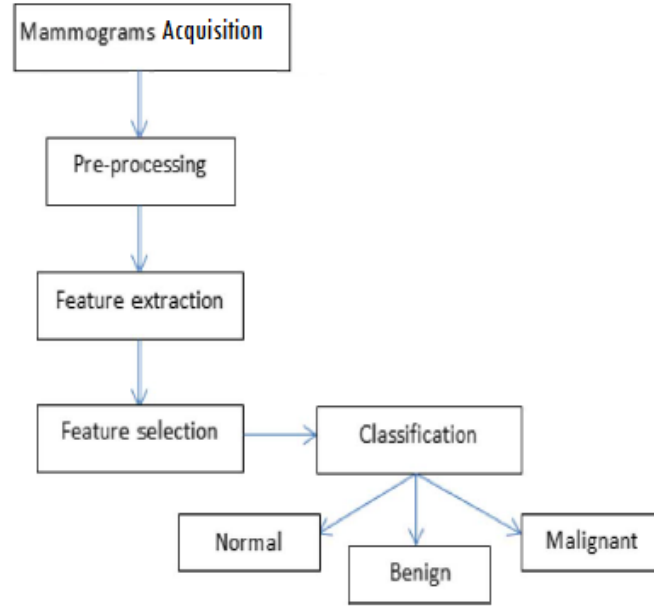


Figure 2.4: Overall structure of CAD cancer detection systems

Figure 2.4. The first stage is image acquisition, which is carried with specialized mammography devices as was shown in Figure 1.1. The output from stage one is a digital image at high spacial resolution. The second stage is the preprocessing stage in which the image is contrast-corrected and non-breast regions such as labels and pectoral tissue are identified and removed. The third stage is the feature extraction stage using different techniques. The output of this stage is a feature vector. Such a set of features may be processed further to remove irrelevant features (Stage 4). Finally, in the classification stage, the test feature vector is used to decide whether the ROI is normal, benign, or malignant .

Berber et al. [28] proposed several classification methods for abnormality detection in digital mammograms for the DDSM database. Different statistical features like Mean, Standard Deviation, Smoothness, Skewness, Energy, and Entropy were all discussed, in addition to LBP texture features. They reported that such features

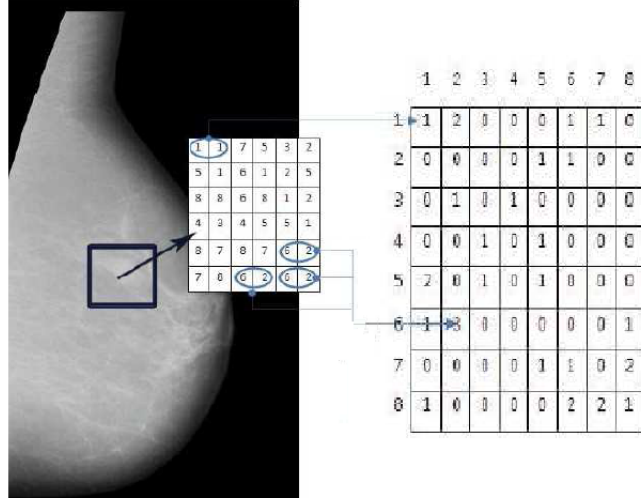


Figure 2.5: An example of GLCM constructed for a small patch from a mammogram image [4].

were very robust for the task of cancer detection. The detection accuracy achieved of 98.63% using an SVM classifier, and 97.25% using a KNN classifier.

Sabu et al. [4] discussed various texture analysis approaches for the detection of masses and micro calcification in mammography. They used GLCM, Gray Level Run Length Matric (GLRLM), LBP, Gray level difference statistics, Laws texture measures and Fractal based texture analysis. An example of the GLCM constructed for a small patch from a given input image is shown in Figure 2.5.

Asad et al. [29] introduced a feature set involving six pre-existing and new features. Thirty-three images from the Mini-MIAS database were selected for training and testing. They included 16 circumscribed benign cases, 4 circumscribed malignant cases, 9 speculated benign cases, and 5 speculated malignant lesions cases. The features were trained using Kohnan Neural Networks (KNN) achieving a 80% classification rate.

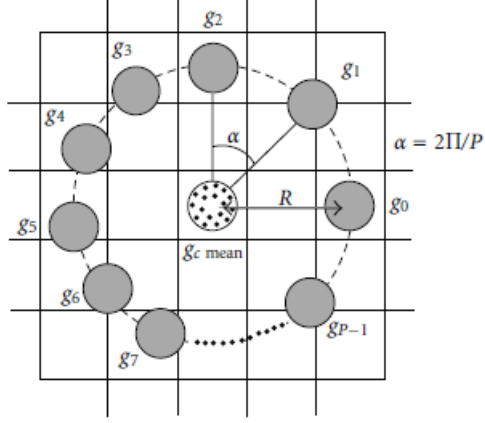


Figure 2.6: The central pixel g_c and its P circularly symmetric neighbors at radius R

Gargouri et al. [30] proposed a local pattern model called Gray Level and Local Difference (GLLD) where absolute gray levels and local differences were used as textural features for mammographic mass detection. The combined descriptors were, respectively, $SGLLD_{24,3}^{riu2}$, $MGLLD_{24,3}^{riu2}$ and $CGLLD_{24,3}^{riu2}$ called GLLD.

$$SGLLD_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_{c\ mean}) 2^p \quad (2.1)$$

Where g_p is the gray level of p -neighbor and P is the total number of neighbors in the radius R that used in processing, and $g_{c\ mean}$ is the average value of the central pixel and $s(x)$ is traditional signum function (see Figure 2.6).

$$s(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (2.2)$$

$$MGLLD_{P,R} = \sum_{p=0}^{P-1} t(m_p, c) 2^p \quad (2.3)$$

$$t(x, c) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases} \quad (2.4)$$

$$\text{CGLLD}_{P.R} = t(g_{c\text{mean}}, c_I) \quad (2.5)$$

The three obtained codes from equations 2.1, 2.3, and 2.5 for each pixel have carry important texture information. The authors estimated the histogram over an image for each code and then concatenated these to build the GLLD feature set. This concatenated histogram was then used in classification. The Artificial neural networks (ANN) classifier was shown to provide excellent performance and can consider different image sizes with improved detection rates. A comparative study with existing local pattern-based features such as LBP showed that the proposed method led to a better performance with a classification rate of 95%.

Oliver et al. [31] proposed a method for mass false positive reduction based on textural features. A global descriptor for each ROI is obtained then a ROI image was divided into small regions and local texture descriptors were then computed using local binary patterns. The final feature vector from combining these local descriptors was used to classify the ROIs as true masses or normal tissues. The authors showed that LBP features are effective and efficient for false positive reduction with different patches sizes.

Luo et al. [32], used two feature selection methods, Forward Selection (FS) and Backward Selection (BS), to remove redundant features for the detection of breast cancer

and then compared the performance between the two methods. The performance of the algorithm was increased by removing the irrelevant features in a similar fashion to density classification. The breast cancer prediction accuracy was determined using DT, SVM-SMO with an enhanced performance. The prediction performance was tremendously increased using the proposed feature selection methods with ensemble classifiers on the breast cancer dataset.

Biswas et al [5], proposed a method for modeling mammogram textures using a mixture of Gaussians distribution (GMM). A two-layer architecture was proposed for classification. The low-level rotation-invariant textural features were collected at different scales. Then, latent textural primitives from features were extracted using the GMM model. The results showed that the proposed probabilistic approach was better than other competing approaches.

The author in [33] focused on the problem of breast density. They collected their own data from Malaysian patients at the University of Malaya Medical Centre (UMMC) during the period 2008 to 2010. The database included standard images of dense, fatty and fatty-glandular breasts, and were classified into three categories: normal, benign and malignant, using the results obtained from biopsies. Malignant and benign abnormalities were selected from the segmented images using the Ground Truth (GT) data and markings obtained from the radiologists. Texture based features were extracted from the ROI samples using GLCMs. For classification, the optimum subset of texture features was used with an SVM classifier. The binary classification

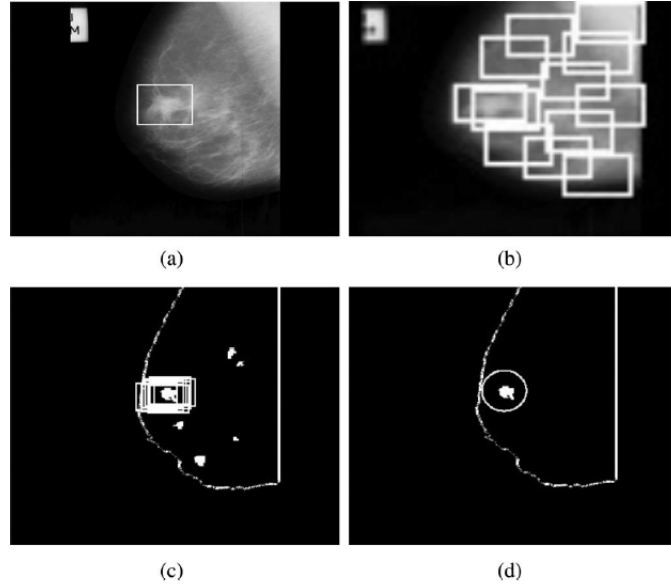


Figure 2.7: Multiscale texture modeling [5] : (a) image marked by a radiologist, (b) Detect abnormal ROIs, (c) Classify ROIs into AD/non-AD, and (d) Detected correct ROI based on (c).

accuracy of the developed system was measured using the Receiver Operating Characteristic (ROC) with performance measures such as sensitivity, specificity and the Area under the Curve (AUC). They also compared their results to a ANN-based classifier and found that the SVM classifier gave better results. Their approach achieved a classification accuracy of 97.6% for classifying images into malignant or benign type.

As seen from the above discussion, cancer detection is the major step in any CAD system. This step is considered to be the main component of any CAD system. Many researchers are still working on developing systems with an acceptable performance from the clinical side. But no comprehensive approach has been adopted to date. The approaches that have been developed for cancer detection focus either on detecting the abnormalities and masses or on classifying cancer into benign and malignant. There are no approaches that detect densities and classify the tissues in a joint fashion.

Also, in the feature extraction step, many approaches were developed using different features focusing on specific types of textures. Here, we develop a framework for finding a feature vector that contains the main features relevant to the task of interest then use these in classification.

2.3 Texture analysis techniques

In most image processing applications, we are interested in the content and the analysis of patterns or regions rather than the specific gray level values of the pixels. In our application, for example, we focus mainly on identifying regions that may that contain cancerous cells rather than the specific pixels themselves. For this reason, among other, the literature is full of comprehensive surveys covering texture analysis topics. In [34], for example, different texture features were discussed and categorized in four groups: statistical features, structural features, signal processing based features, and model based features. They considered the co-occurrence matrix as one of the most popular techniques for estimating different second order statistics. They also considered the Gray Level Run Length (GLRL) and the LBP as higher order statistical features. Finally, they categorized the wavelet transform based features and Gabor analysis as signal processing based features. As many of these features are used in our work, we will briefly discuss each one of these with some insights on their advantages and disadvantages.

2.3.1 Local Binary Patterns (LBP) texture features

Ojala and et al. [6] introduced a new multi-resolution approach for rotation invariant texture features named local binary patterns. Their approach was initially developed for gray scale images and used to detect 'uniform' patterns at any spatial resolution:

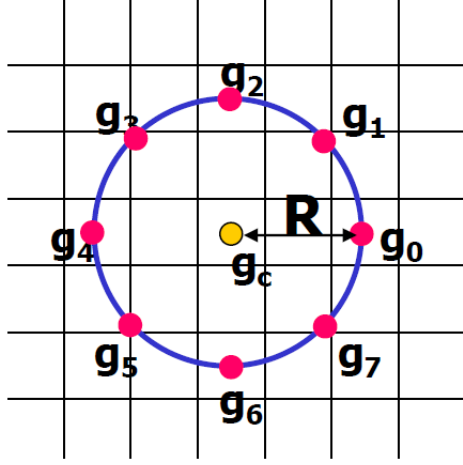


Figure 2.8: LBP for Circular neighborhood [6]

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (2.6)$$

$$s(g_x - g_y) = \begin{cases} 1, & (g_x - g_y) \geq 0 \\ 0, & (g_x - g_y) < 0 \end{cases} \quad (2.7)$$

Where $LBP_{P,R}$ is the traditional LBP texture at g_c (central pixel value) which is modeled using a local neighborhood of radius R , and sampled at P points as shown in Figure 2.8. The $LBP_{P,R}$ provides the relation between a pixel with neighbors where P bits assigned for each pixel and if the value of gray level of the neighbor greater than a pixel value, then the bit belongs to that neighbor in the pattern will be 1 otherwise

0. An example of the traditional approach is shown in Figure 2.9.

Many improvements over the traditional version from $LBP_{P,R}$ have been developed. One of these versions is Local Binary Pattern uniform and rotational-invariant ($LBP_{P,R}^{riu2}$). $LBP_{P,R}^{riu2}$ was developed for calculating the LBP but taking into account the uniformity feature with rotation invariance for the LBP pattern. The $LBP_{P,R}^{riu2}$ is obtained using the following equation:

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(LBP_{P,R}) \leq 2 \\ P + 1 & \text{otherwise} \end{cases} \quad (2.8)$$

Where $LBP_{P,R}^{riu2}$ stands for LBP uniform and rotational invariant with P samples located in radius R. $U(LBP_{P,R})$ is calculated as follows

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (2.9)$$

Where the uniform pattern is defined as bit patterns with 0 or 2 transitions 01 or 10.

This approach was shown to be robust against gray-scale variations since the operator was, by definition, invariant against any monotonic transformation of the gray scale. For feature extraction purposes, a histogram is obtained from the $LBP_{P,R}^{riu2}$ image and used as a feature vector for the classification stage. As an example, the histogram contains 10 bins when the number of samples $P = 8$. The first nine bins come from the count of ones in the pattern where starting with 0 when no ones exist and 8 when all pattern is ones. The tenth bin is considered for non-uniform LBP pattern.

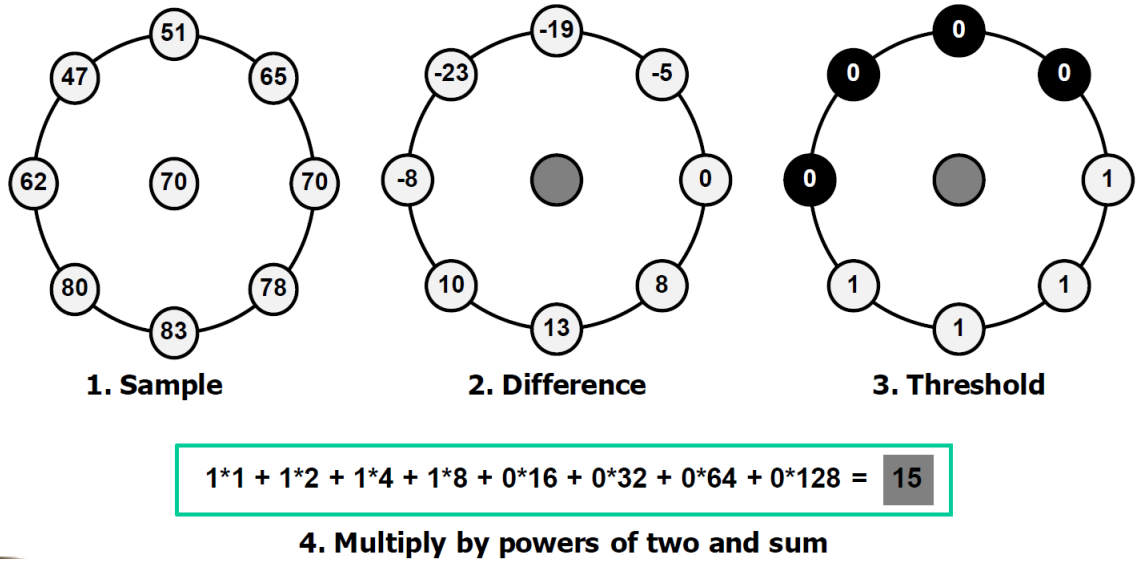


Figure 2.9: An example of finding LBP with $R=1, P=8$ and gray level value $g_c = 70$

2.3.2 Gray Level Co-occurrence Matrix (GLCM) features

The spatial GLCM was originally introduced by Haralick et. al in [35]. The GLCM calculates the occurrence of the gray levels with neighbors in different angles(θ) such as $0^\circ, 45^\circ, 90^\circ$ and 135° (See Figure 2.10) and the pair placed in distance d where d here equal 1. The distance and angles are the main parameters for calculating the GLCM as the occurrences will be counted based on the distance between the pixels and in a given direction [35, 36, 37]. From the example shown in Figure 2.10, $P(1,1)$ is the occurrence of the pixel with gray level equal 1 followed by gray level equal 1 in the same direction (angle = 0°) and the number of occurrences equals 1. We list in Table 2.5, the main notation used in estimating the different features for the GLCM. These features are now briefly discussed.

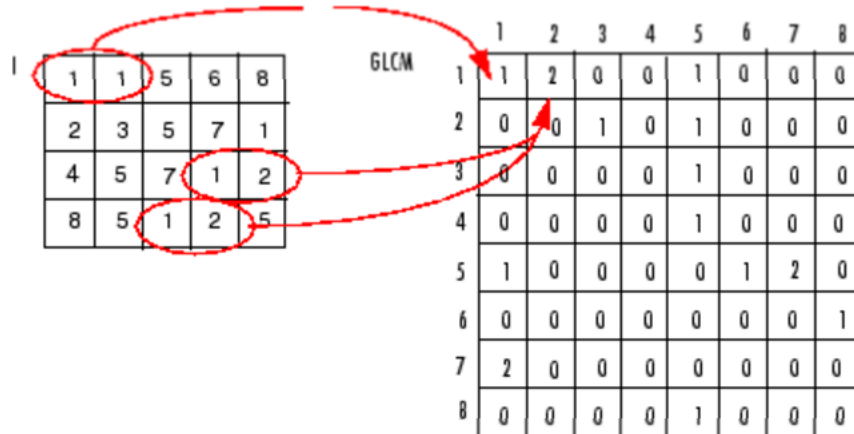


Figure 2.10: An example of calculating GLCM for 0°

Autocorrelation Feature

The autocorrelation feature is based on the repetitive nature of textures observed within a given image. The correlation between the image and its translated version is measured using this feature.

$$Autocorrelation = \sum_i \sum_j (ij)p(i, j) \quad (2.10)$$

Contrast Feature

The spatial frequency is measured using the contrast feature computed from the difference between highest and lowest values for a set of contiguous pixels. Mainly, it measures the amount of local variations in the image. So, for low contrast images, the values of GLCM are concentrated around the low spatial frequencies and the main

Table 2.5: Basic notation used in GLCM computations

$p(i, j)$	(i,j) entry of the co-occurrence probability matrix.
L	Number of Gray-levels.
v	Mean value of $p(i, j)$
$p_x(i)$	$\sum_{j=1}^L p(i, j)$
$p_y(i)$	$\sum_{i=1}^L p(i, j)$
μ_x	$\sum_i \sum_j i \cdot p(i, j)$
μ_y	$\sum_i \sum_j j \cdot p(i, j)$
σ_x^2	$\sum_i \sum_j (i - \mu_x)^2 \cdot p(i, j)$
σ_y^2	$\sum_i \sum_j (j - \mu_y)^2 \cdot p(i, j)$
$p_{x+y}(k)$	$\sum_{i=1}^L \sum_{j=1}^L p(i, j), \quad k = 2, 3, \dots, 2L$ $i+j=k$
$p_{x-y}(k)$	$\sum_{i=1}^L \sum_{j=1}^L p(i, j), \quad k = 0, 1, \dots, L - 1$ $ i-j =k$
HX	$-\sum_i p_x(i) \cdot \log(p_x(i))$
HY	$-\sum_i p_y(i) \cdot \log(p_y(i))$
HXY	$-\sum_i \sum_j p(i, j) \cdot \log(p(i, j))$
$HXY1$	$-\sum_i \sum_j p(i, j) \cdot \log(p_x(i) \cdot p_y(j))$
$HXY2$	$-\sum_i \sum_j p_x(i) \cdot p_y(j) \cdot \log(p_x(i) p_y(j))$

diagonal.

$$Contrast = \sum_{n=1}^{L-1} n^2 \left\{ \sum_i \sum_j p(i, j) \right\}, |i - j| = n, L = \text{number of gray levels} \quad (2.11)$$

Correlation Feature

This feature measures the linear dependency between gray levels in the image.

$$Correlation = \frac{\sum_i \sum_j (ij)p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (2.12)$$

Cluster Prominence Feature

The cluster prominence measures the asymmetry where the image with less symmetric yields high value of cluster prominence and vice versa. A peak in the GLCM around the mean value exists when the cluster prominence value is low.

$$ClusterProminence = \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i, j) \quad (2.13)$$

Cluster Shade Feature

The cluster shade measures the asymmetry too, where it measures the skewness of the GLCM distribution.

$$ClusterShade = \sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i, j) \quad (2.14)$$

Dissimilarity Feature

The dissimilarity feature measures the amount of differences between two pixels.

$$Dissimilarity = \sum_i \sum_j |i - j| p(i, j) \quad (2.15)$$

The GLCM Based Energy Feature

This feature measures texture uniformity . When the gray distribution has constant or periodic structure form then the value of energy will be high which is related to

the homogeneity of the image. The maximum value of energy equals one.

$$Energy = \sum_i \sum_j p(i, j)^2 \quad (2.16)$$

Entropy Feature

This feature measures the complexity of an image. When the GLCM values follow a nonuniform distribution, then the entropy value will be high.

$$Entropy = - \sum_i \sum_j p(i, j) \cdot \log(p(i, j)) \quad (2.17)$$

Homogeneity Feature

This feature measures the gray tone differences between pair elements. It reaches its the maximum value when the image has identical gray level values. The homogeneity and contrast have a strong relationship since when the homogeneity increases, the contrast feature decreases, while the energy can be kept constant. It is highly sensitive and related to the diagonal of the GLCM.

$$Homogeneity = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j) \quad (2.18)$$

Maximum Probability feature

This feature shows the pair of gray-levels occurring more than the other pairs in certain direction and at a certain distance.

$$\text{Maximumprobability} = \underset{i,j}{MAX} p(i, j) \quad (2.19)$$

Variance feature

This feature increases when the gray level values differ more from the mean of the image.

$$\text{Variance} = \sum_i \sum_j (i - v)^2 p(i, j) \quad (2.20)$$

Sum average feature

$$\text{Sumaverage} = \sum_{i=2}^{2L} i \cdot p_{x+y}(i), i = 1, 2, \dots, L \text{ and } j = 1, 2, \dots, L \quad (2.21)$$

Sum variance feature

$$\text{Sumvariance} = \sum_{i=2}^{2L} (i - \text{Sumentropy})^2 \cdot p_{x+y}(i), 2L : i = 1, \dots, L \text{ and } j = 1, \dots, L \quad (2.22)$$

Sum entropy feature

$$\text{Sumentropy} = - \sum_{i=2}^{2L} p_{x+y}(i) \cdot \log(p_{x+y}(i)) \quad (2.23)$$

Difference variance feature

$$\text{Differencevariance} = \sum_{i=0}^{L-1} i^2 \cdot p_{x-y}(i) \quad (2.24)$$

Difference entropy feature

$$Differenceentropy = - \sum_{i=0}^{L-1} p_{x-y}(i) \cdot \log(p_{x-y}(i)) \quad (2.25)$$

Information measure of correlation 1

$$Informationmeasureofcorrelation1 = \frac{HXY - HXY1}{\max(HX, HY)} \quad (2.26)$$

Information measure of correlation 2

$$Informationmeasureofcorrelation2 = \sqrt{1 - \exp[-2(HXY2 - HXY)]} \quad (2.27)$$

Inverse difference normalized (INN) feature

$$Inversedifferencenormalized(INN) = \sum_i \sum_j \frac{p(i, j)}{1 + |i - j|^2/L} \quad (2.28)$$

Inverse difference moment normalized feature

$$Inversedifferencemomentnormalized = \sum_i \sum_j \frac{p(i, j)}{1 + (i - j)^2/L} \quad (2.29)$$

2.3.3 Gray Level Run Length Matrix (GLRL) features

Another framework for extracting texture feature was developed by Galloway [38] which he called Gray Level Run Lengths (GLRL). A gray level run is a set of consecutive, collinear pixels that have the same gray level values. The length of the run is the number of pixels in the run. For a given image, the GLRL can be computed for runs in any given direction. In the given direction, the matrix element (i,j) specifies the number of times that the image contains a run of length j of pixels value i. The following features are extracted from the GLRL textures.

Let $q(i, j)$ be (i,j) entry of the $M \times N$ run-length matrix, and

n_r is the total number of runs,

n_p is the number of pixels in the image

Short Run Emphasis (SRE) feature

This feature emphasizes short runs.

$$SRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{q(i, j)}{j^2} \quad (2.30)$$

Long Run Emphasis (LRE) feature

This feature emphasizes long runs.

$$LRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N q(i, j) \cdot j^2 \quad (2.31)$$

Gray-Level Non-uniformity (GLN) feature

This feature measures the gray level non uniformity of the image. When runs are uniformly distributed, the GLN feature results in low values.

$$GLN = \frac{1}{n_r} \sum_{i=1}^M \left(\sum_{j=1}^N q(i, j) \right)^2 \quad (2.32)$$

Run Length Non-uniformity (RLN) feature

This function measures the non-uniformity of run lengths where the runs are uniformly distributed during the lengths, the RLN value decreases.

$$RLN = \frac{1}{n_r} \sum_{j=1}^N \left(\sum_{i=1}^M q(i, j) \right)^2 \quad (2.33)$$

Run Percentage (RP) feature

This function is a ratio of the total number of runs to the total number of possible runs if all runs had a length of one. For linear-most patterns, the RP results in the lowest value.

$$RP = \frac{n_r}{n_p} \quad (2.34)$$

Low Gray-Level Run Emphasis (LGRE) feature

$$LGRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{q(i, j)}{i^2} \quad (2.35)$$

High Gray-Level Run Emphasis (HGRE) feature

$$HGRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N q(i, j) \cdot i^2 \quad (2.36)$$

Short Run Low Gray-Level Emphasis (SRLGE) feature

$$SRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{q(i, j)}{i^2 \cdot j^2} \quad (2.37)$$

Short Run High Gray-Level Emphasis (SRHGE) feature

$$SRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{q(i, j) \cdot i^2}{j^2} \quad (2.38)$$

Long Run Low Gray-Level Emphasis (LRLGE) feature

$$LRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{q(i, j) \cdot j^2}{i^2} \quad (2.39)$$

Long Run High Gray-Level Emphasis (LRHGE) feature

$$LRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N q(i, j) \cdot i^2 \cdot j^2 \quad (2.40)$$

2.3.4 Wavelet based texture analysis

The introduction of wavelets in different signal and image applications prompted many researchers to formulate features based on such a transform. Wavelet based texture analysis uses a class of functions, that are localized in both spatial and spatial-frequency domains, to decompose texture images. Wavelet functions are constructed from a basis function called 'mother wavelet' or 'basic wavelet', by means of dilation and translation [39]. For image processing, the 2D wavelet transform is characterized by 2 features: the scaling and wavelet functions denoted as $\phi(x, y)$ and $\psi(x, y)$, respectively. The scaled and translated basis functions are defined as:

$$\begin{aligned} \phi_{j,m,n}(x, y) &= 2^{j/2} \phi(2^j x - m, 2^j y - n), \\ \psi_{j,m,n}^i(x, y) &= 2^{j/2} \psi^i(2^j x - m, 2^j y - n), i = \{H, V, D\} \end{aligned} \quad (2.41)$$

where index i defines the direction of the wavelet functions such that $\psi_{j,m,n}^H(x, y)$ measures variations along columns, $\psi_{j,m,n}^V(x, y)$ measures variations along rows, and $\psi_{j,m,n}^D(x, y)$ measures variations along diagonals. Here, we get three different wavelet functions based on the translation of the basis function for each level. Using the

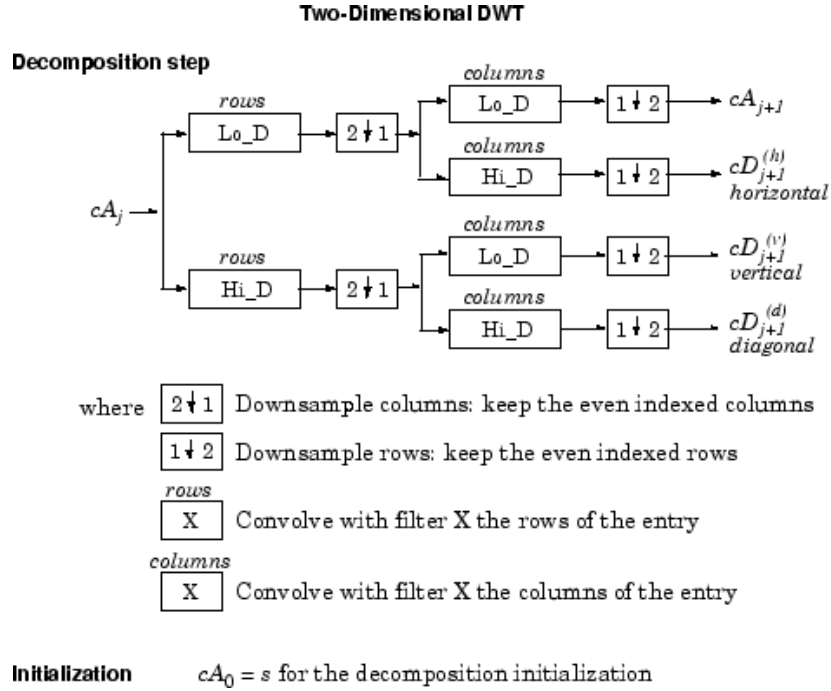


Figure 2.11: The structure of a 2D Wavelet decomposition with 1 Level [7]

traditional wavelet decomposition [8], images are decomposed at each level into 4 sub images as shown in figure 2.13. The structure of a 2D wavelet decomposition for 1 and 2 levels is shown in Figure 2.11 and 2.12. The lower resolution images denoted as LL, where LH, HL and HH images result from the wavelets decomposition using ψ^H , ψ^V , and ψ^D respectively. The LL image can further be decomposed into 4 new sub images, as so on. Many features can be extracted from the sub images such as mean, variance, energy, entropy,...etc. These features are useful in many applications including compression, classification, etc.

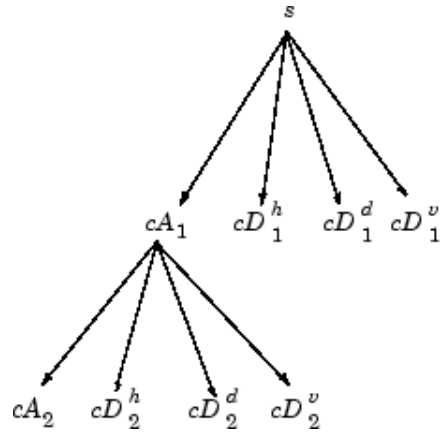


Figure 2.12: The structure of a 2D Wavelet decomposition with 2 Levels [8]

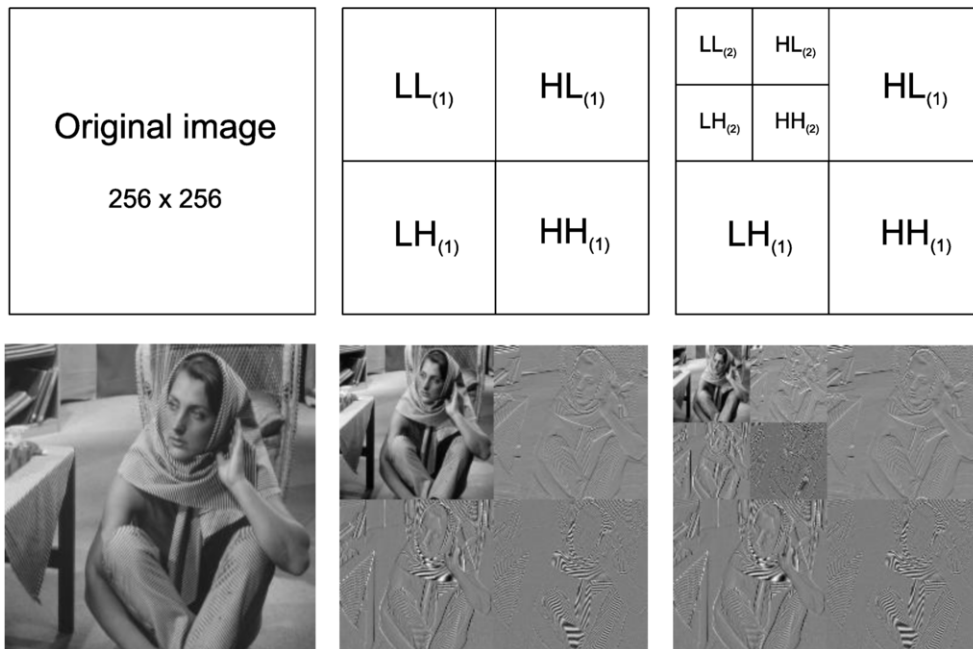


Figure 2.13: An example for 2D Wavelets decomposition with 2 Levels

2.3.5 Gabor texture analysis

Daugman [40] originally proposed a filter-based approach using neural networks. The characteristics of the Gabor wavelets (filters), especially for frequency and orientation representations, are similar to those of the human visual system. As one of the applications of the Gabor filters, they are found to be particularly appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave [41]

$$\begin{aligned} G(x, y) &= \frac{f^2}{\pi\gamma\eta} \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp(j2\pi f x' + \phi) \\ x' &= x \cos \theta + y \sin \theta \\ y' &= -x \sin \theta + y \cos \theta \end{aligned} \tag{2.42}$$

where f is the frequency of the sinusoidal factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function, ϕ is the phase offset, σ is the standard deviation of the Gaussian envelope and γ is the spatial aspect ratio which specifies the ellipticity of the support of the Gabor function.

Gabor filters are used to model the spatial summation properties of simple cells in the visual cortex and have been adapted and popularly used in texture analysis. They have been long considered as one of the most effective filtering techniques to extract useful texture features at different orientations and scales. Gabor filters can be categorized into two components: a real part as the symmetric component and an imaginary part as the asymmetric component. An example of Gabor wavelets over five scales and eight orientations is shown in Figure 2.14, where the energy and entropy can be extracted

from each sub images as features of the Gabor filter. There are many advantages of using Gabor filter the detection of variations in different angles including and scales. Hover, the transformation is complex and hence computationally expensive.

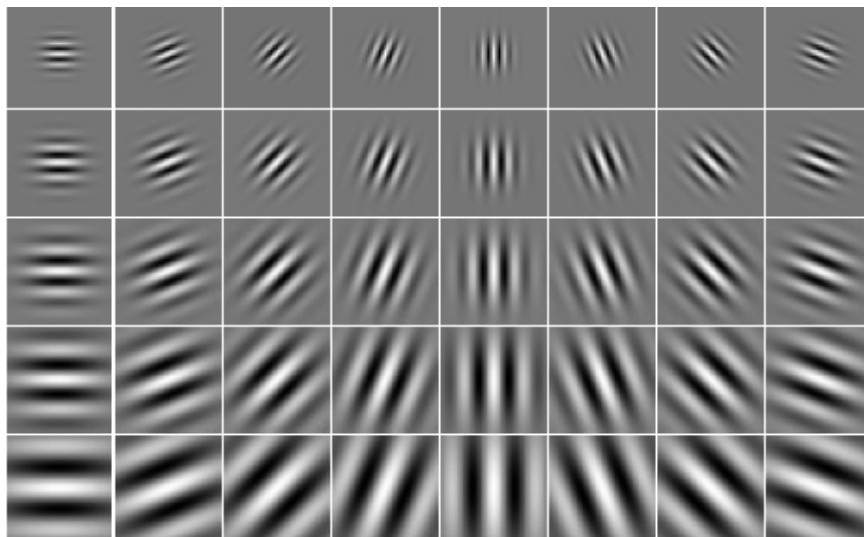


Figure 2.14: An example of Gabor filter masks over five scales and eight orientations

2.3.6 Summary

In this chapter, we briefly discussed the literature related to feature extraction for image which is particularly useful in density classification from mammogram images. In density classification, we presented the benchmark in density classes used for density classification. After that, many approaches for cancer detection were discussed in the section 2.2. Also, we discussed the accuracy rate for these approaches, their advantages and disadvantages. In the last section (2.3), we discussed some popular texture approaches used in divers applications. Next, we will implement our approach for feature extraction followed by the classification algorithm.

CHAPTER 3

RESEARCH METHODOLOGY

In this chapter, we present our proposed approach for both density and cancer classification. We discuss the implementation of the different stages of our system (Figure 3.1). We discuss how the system was tested with the MIAS database and the Image Retrieval in Medical Applications (IRMA) patch database.

3.1 Proposed approach

Our proposed approach follows the general structure of traditional CAD systems mentioned in chapter 2 with an important improvement. Our work is divided into two main parts. In the first part, we analyze full 1024×1024 mammogram images and apply the preprocessing stage only over these. In the second part, we assume that the ROI patches of size 128×128 have already been extracted from the IRMA database. In this part, an optimal set of features is obtained and used to perform the proposed 2-stage classification for both density classification and cancer detection. A complete block diagram for our proposed system is shown in Figure 3.1. We will now discuss

each of the stages of the proposed CAD system.

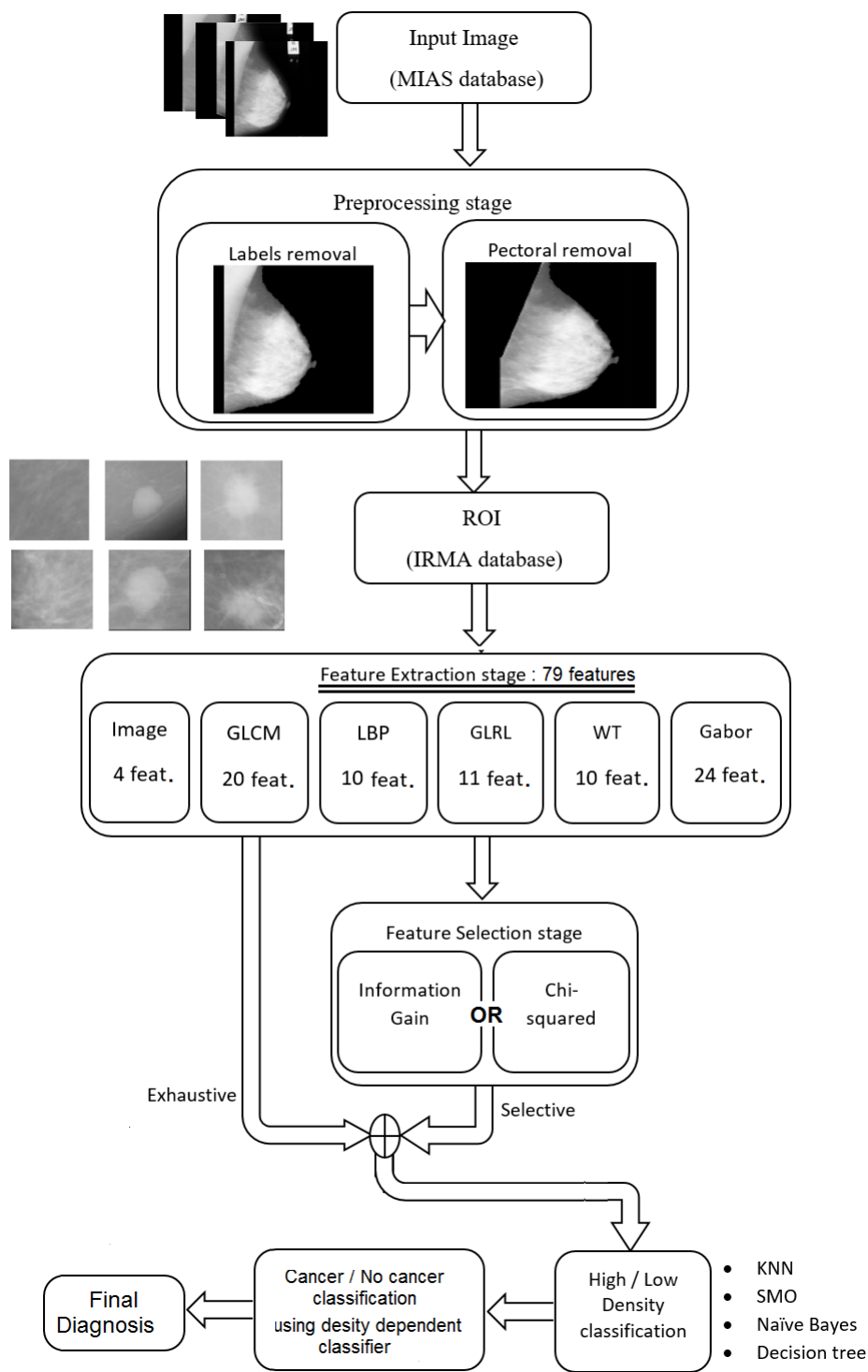


Figure 3.1: Full view of the proposed CAD system

3.1.1 Image Preprocessing

The main aim of this stage is to extract the breast area only as the original mammogram contains many objects with the breast area as was shown in Figure 1.2. These include the pectoral muscle and labels, so on. So, we need to remove all unnecessary objects for the mammogram images. Many researchers discussed diverse techniques of pectoral removal [9][42]. An example of applying these common steps is shown in Figures 3.2 and 3.3.

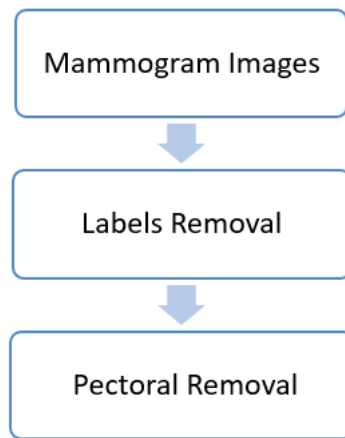


Figure 3.2: Preprocessing steps

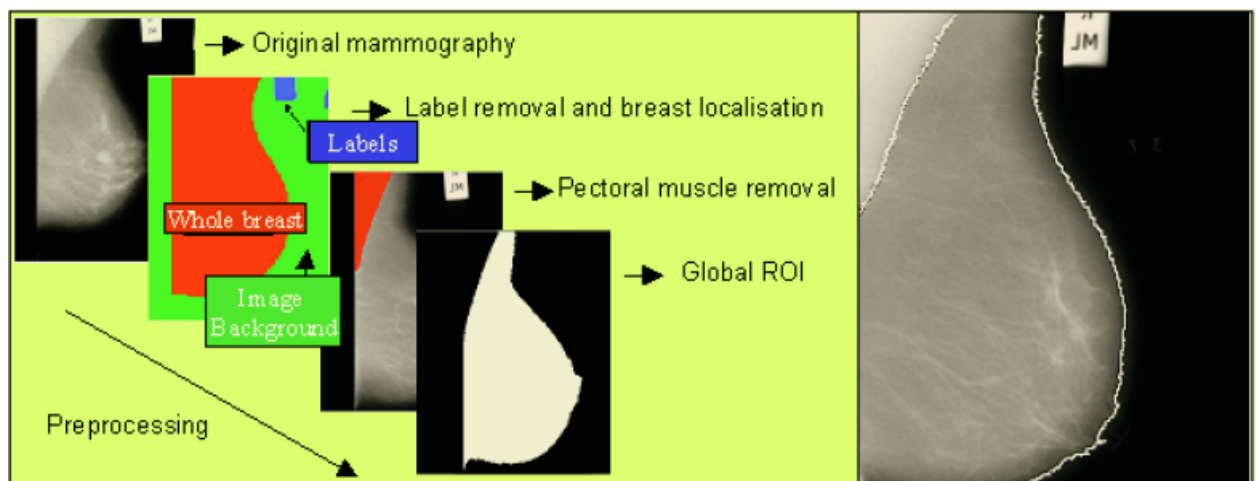


Figure 3.3: An example of applying a sequence of preprocessing steps [9]

Here, to remove the labels, we extract the objects in the image using a basic thresholding algorithm from the original image e.g. Figure 3.4-(a) , then keep the largest object and remove the others. The resulting image contains the breast region with the pectoral as shown in Figure 3.4-(b).

At this point, we have the breast and pectoral regions as an output from the previous step. We start by applying the Otsu's thresholding algorithm [43] on the breast region with some basic features from image processing tools on MATLAB [44, 45]. Using region features, we determine the location of pectoral area in the image. Then, a region growing technique is applied to extract the pectoral area and keep only the breast area as shown in Figure 3.4-(c).

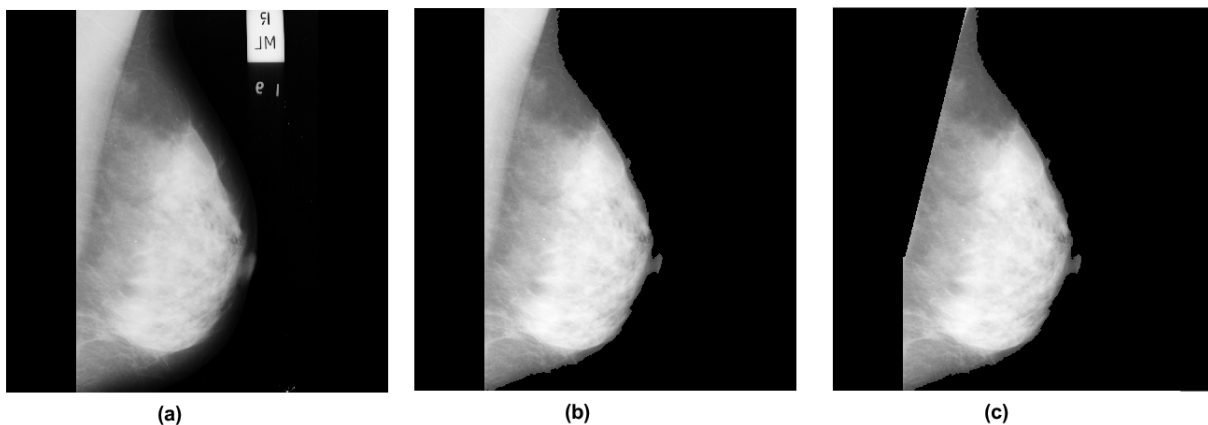


Figure 3.4: An example of applying a sequence of preprocessing steps: (a) the original image, (b) after applying label removal, and (c) after applying pectoral removal

3.1.2 Feature extraction

Among different feature extraction approaches listed in [34], some of these are used individually or in combination with each other to improve classification accuracy. In

the previous section of a texture analysis, we discussed a number of feature extraction approaches that are commonly used in mammogram analysis: LBP , GLCM and wavelet transform features, etc. . In our work, we started by computing a galaxy of 79 features distributed as following:

- **Statistical features:** These are 4 features (Mean, Variance, Kurtosis, and Skewness) which are extracted from the images/patches directly.
- **GLCM features:** These are 20 features which were discussed in section 2.3.2.
- **GLRL features:** These are 11 features discussed earlier in section 2.3.3.
- **LBP features:** These are 10 features which are generated by obtaining the histogram of LBP values which were discussed in section 2.3.1.
- **Wavelet transform features:** These are 10 features based on applying a multi-level wavelet transform with 3 levels. The number of levels is chosen based on the experiments carried before showing no additional information or improving in classification accuracy is achieved when we increase the number of levels further. So, we will have 3 bands for each level except for the third level we have 4 bands which we discussed in section 2.3.4. We calculate the percentage of energy corresponding to each band (the approximation, horizontal, vertical and diagonal details images) .
- **Gabor filter features:** These are 24 features where we used 3 scales with 8 orientations. Also, no more information or improving in classification accuracy

was achieved when we increased the number of scales or orientations further.

So, we have 24 bands, we, then, calculate the energy for each band.

After calculating the different features, we combine these in one super feature vector which is then used in classification. The feature vector can also be reduced using optimal feature selection. The above listed features are used in our proposed approach throughout our experiments.

3.1.3 Feature selection

We mentioned above that the size of the original feature vector can be reduced by removing irrelevant features or features with low correlation with class labels. In the feature selection stage, we identify the optimal features to be used in the classification stage. The optimal features are identified by ranking the features and using the top ranked ones. The main tool that we used in this work is the WEKA software [46][47] which includes a collection of machine learning algorithms and data processing tools.

The algorithms we tested here are: the Chi-Squared statistic Evaluator and the Information Gain Evaluator [48]. For the Chi-Squared statistics, the algorithm evaluates the importance of features by computing the value of the Chi-squared taking into account the class labels. The Chi-squared value is obtained from the sample variance or sum of squared errors. For the Information Gain(IG) algorithm, the importance of the feature is obtained by measuring the IG with respect to the class labels:

$$IG(class, feature) = H(class) - H(class|feature) \quad (3.1)$$

Which measures the entropy when this feature is considered or not.

In our work, we tried both scenarios: with and without optimal feature selection. This stage is followed with the classification stage either for density classification or for cancer detection.

3.1.4 Pattern Classification

This stage is divided into two phases: breast density classification and cancer classification in sequential manner. To fully investigate the power of the proposed algorithm, we tested four classifiers. These are :

- Support Vector Machine based on Sequential Minimal Optimization training (SMO-based on SVM)[49][50][51]: The SVM classifier is one of the most popular classification algorithms based on learning using supervised approaches. The training itself of the SVM classifier can be carried in a number of ways. Here, we used an SVM implementation based on Sequential Minimal Optimization (SMO). The SMO algorithm is used for solving quadratic programming (QP) problems useful in training SVMs. The SVM classifier used here was implemented using a polynomial kernel:

$$K(x, y) = \langle x, y \rangle^p$$

where x and y are vectors in the input space. This classifier was shown to give best results in the classification stage as will be seen later.

- the KNN classifier [52][53]: The KNN classifier is a supervised learning algorithm and considers instance-based classification. The KNN operates on the premises that classification of unknown instances can be achieved by relating the unknown pattern to the known ones based on a given similarity measure (e.g., distance functions). This classifier used the k-nearest neighbor principle to determine the class using Euclidean distance.
- Decision tree classifier [54]: The Decision Tree Classifier is a non-parametric supervised learning method used for classification. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. Tree models in which target variables can take a finite set of values are called classification trees. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to these class labels. The best results achieved when the attributes or features are of categorical nature [55].
- Naive Bayes classifier[56]: The Naive Bayes classification covers a set of supervised learning algorithms based on applying the Bayes theorem with the naive assumption of independence between pairs of features. Given a class variable y and a dependent feature vector x_1, \dots, x_n , the Bayes theorem states the following relationship:

$$P(y|x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n|y)}{P(x_1, \dots, x_n)} \quad (3.2)$$

Using the variable independence assumption, we can write:

$$P(x_i|y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i|y) \quad (3.3)$$

Since $P(x_1, \dots, x_n)$ is constant given the input, we can use the following classification rule:

$$\begin{aligned} P(y|x_1, \dots, x_n) &\propto P(y) \prod_{i=1}^n P(x_i|y) \\ \Rightarrow \tilde{y} &= \arg \max_y P(y) \prod_{i=1}^n P(x_i|y) \end{aligned} \quad (3.4)$$

In [57], an empirical study of this classifier was discussed with an explanation as to why this classifier is popular and powerful for diverse applications.

Breast density classification

Density classification is an important step in cancer detection systems. The literature showed that tumor is a mass looking like dense tissue. The classification of density has been discussed by many researchers under different standards such as BIRADS, Low and High densities, etc.

In our work, we considered two cases: Four classes BIRADS, and low (BIRAD I) and high (BIRAD IV) densities. These are shown in Figure 3.5 .

Cancer classification

In cancer classification, the setup considered in this work is shown in Figure 3.6. Normal is equivalent to no cancer in the patch. Benign is the case when a mass exists but is not considered to be cancer by medical practitioners. Malignant represents the

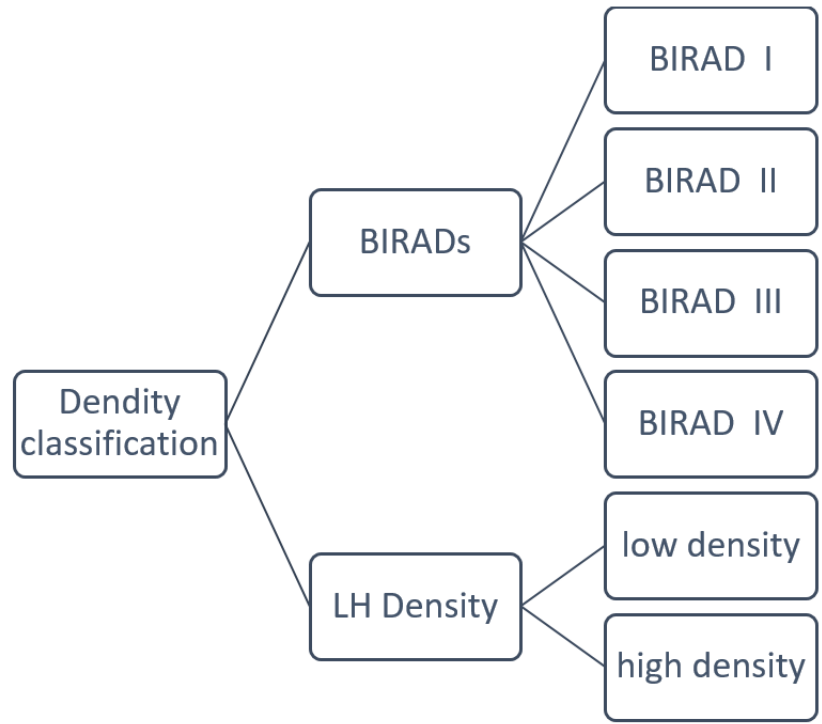


Figure 3.5: Density classification categories

case of a tumor that needs to be removed.

There is also a second type of classification which considers patches as either normal or cancerous where both benign and malignant cases are considered under one cancer class.

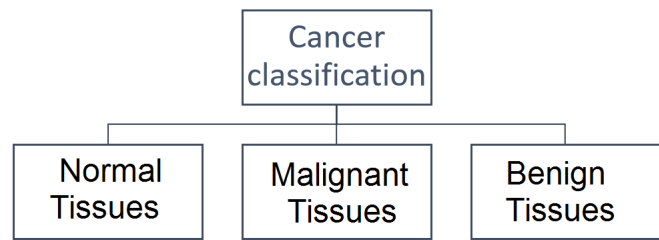


Figure 3.6: Cancer classification categories

3.2 System Implementation

The implementation of our proposed method is discussed in this section where we start with a description of the databases we used, then we discuss the different experimental setups that we considered.

3.2.1 Image Databases

We used two main public databases; the MIAS and the IRMA databases.

MIAS database

This database [58] was generated by MIAS, a UK organization of research groups interested in mammograms. The images were digitized at $50\mu/pixel$ where each pixel is 8 bits. These images have been reduced to $200\mu/pixel$ and each image is 1024×1024 pixels. The database contains 322 images where 106 images are fatty, 104 fatty-glandular and 112 dense-glandular. This database was marked by radiologists and classified into three density classes : fatty, fatty-glandular and dense-glandular. The abnormal images have been classified into either benign and malignant. Figure 3.7 shows three sample images with different densities.

IRMA patch database

The IRMA patches [59] were extracted from different databases like the DDSM, the MIAS, the Lawrence Livermore National Laboratories (LLNL), and the RWTH databases. Each patch has a resolution of $50\mu/pixel$ and 128×128 pixels. This col-

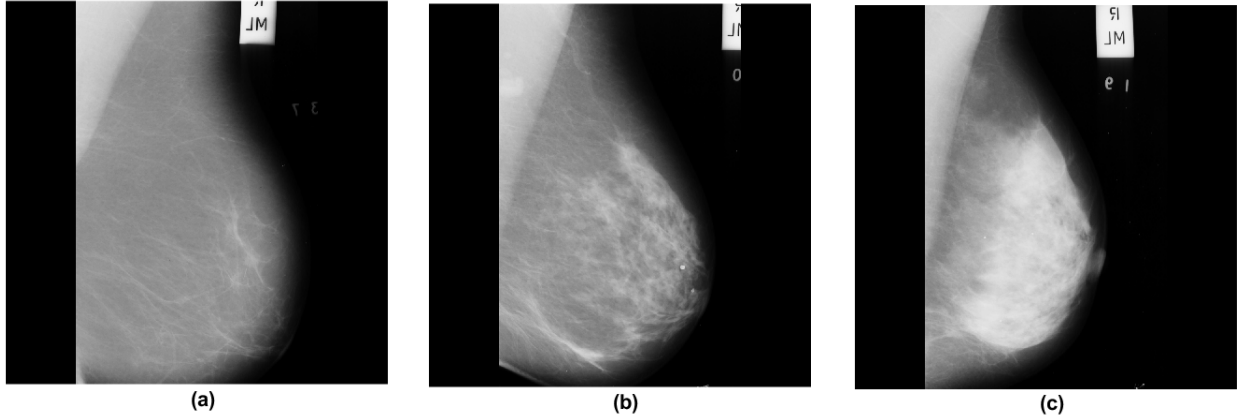


Figure 3.7: Examples of the MIAS database: (a) Fatty - mdb006, (b) Fatty-glandular - mdb008, (c) Dense-glandular - mdb004

lection contains of 12 classes covering 4 BIRADS classes and for each BIRAD class, there are 3 types of patches normal, benign, and malignant. So, we have a total of 12 classes and each class contains of 233 patches. All together, we have 2796 patches with 12 classes. Figure 3.8 shows some examples from the 12 classes patches. In the following table, some statistics about the IRMA dataset are listed.

Description	No. of images	Total images
MLO patch images	1471	2796
CC patch images	1325	
Normal patch images	932	2796
Benign patch images	932	
Malignant patch images	932	
Left breast patch images	1337	2796
Right breast patch images	1459	

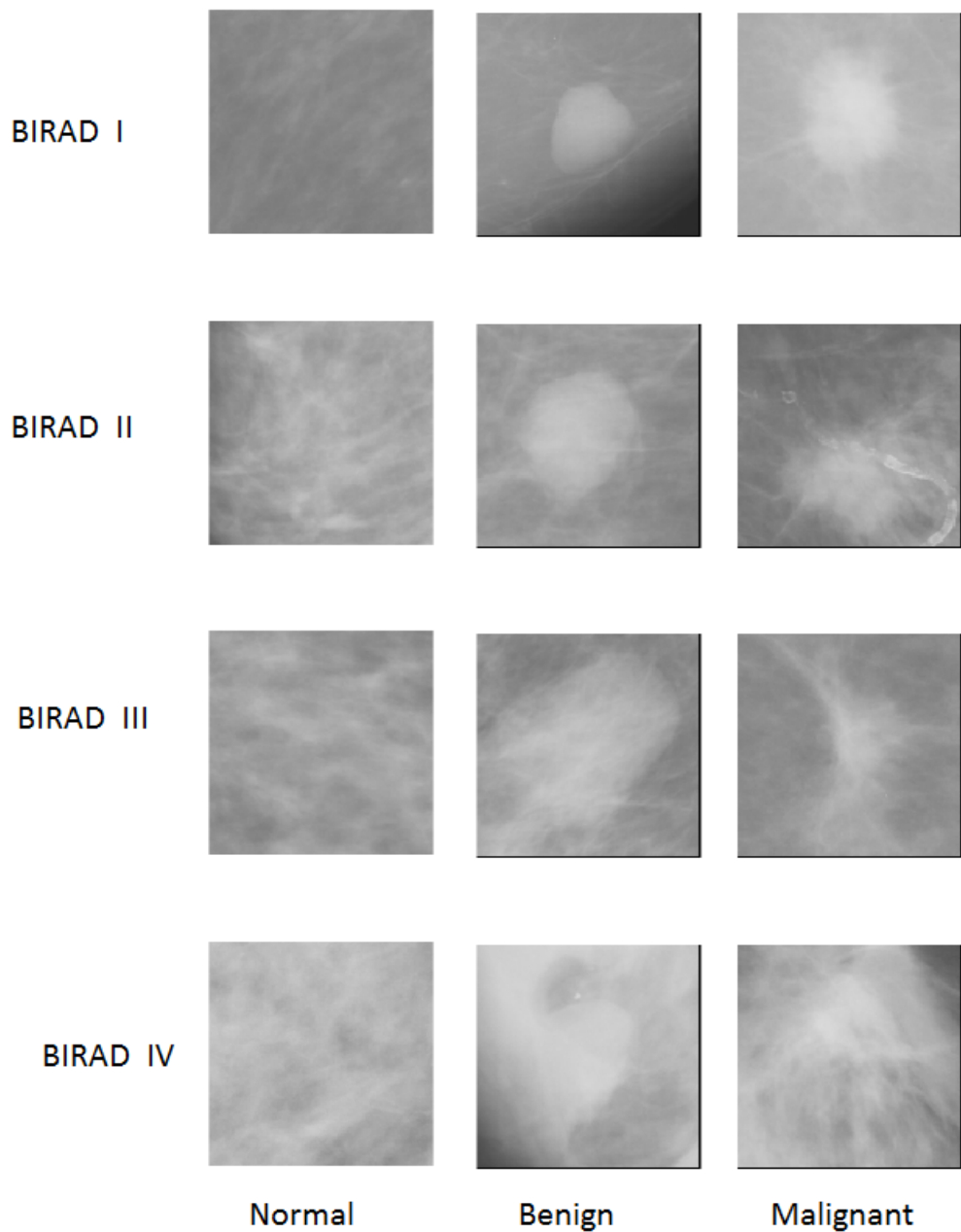


Figure 3.8: Samples from the IRMA database: Each row shows a given BIRAD type

3.2.2 Experimental setup

This section describes the experiments that we carried where we apply the proposed method under different scenarios. In the first experiment, we discuss the results

from the preprocessing stage over full mammograms, then we discuss three different experiments using the patches.

Experiment 1: Pectoral removal of mammography

In this experiment, our algorithm for pectoral removal is discussed (see section 3.1.1). The algorithm starts with an input image from the MIAS database. Then, label and pectoral removal steps are applied as discussed earlier. The input image contains labels and pectoral regions as extra unwanted areas as shown in Figure 3.9-(a). In first step, label removal is applied on the input image and the resulted image is shown in Figure 3.9-(b). In the second step, the pectoral area is removed to get the output image as shown in Figure 3.9-(c). The output image is used for further processing.

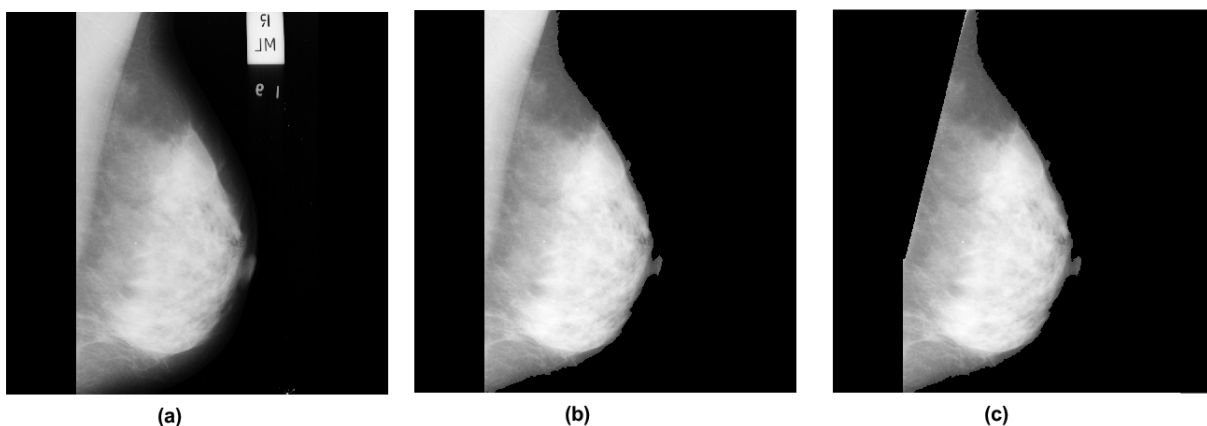


Figure 3.9: An example of applying the preprocessing steps on the mdb004 image: (a)the original image, (b) after applying labels removal, and (c) after applying pectoral removal

Experiment 2: Cancer detection knowing the BIRADS class

In this experiment, we assume the BIRADS type to be known as a prior-knowledge. The experiment performed in the following steps: an input image is taken from the IRMA database where 100 images were chosen randomly from each class. Then, an optimal set of features is obtained using features extraction and selection techniques discussed earlier. Finally, a suitable classifier is obtained for each class. We performed the experiments under different scenarios. For the first setup, we deal with one of these three cancer types: normal, benign, or malignant. Two scenarios were developed for this setup. The two scenarios are: classification using 4 BIRADS classes and classification using Low and High densities. Under each scenario, we obtained cancer classification rate for each scenario individually. For the second setup, we consider the existence of a tumor or not. We also consider the same two scenarios as above. The main assumption for these experiments is that breast density is already known before cancer detection.

Experiment 3: Cancer detection without knowing BIRADS

Here, we ignore the density information and focus only on cancer classification. We followed the same steps discussed above. The input images are taken from the datasets then, an optimal set of features obtained using features extraction and selection techniques. Finally, a suitable classifier is used for each class. Here, we also consider two setups. 100 images were chosen randomly for each class. We also considered here two scenarios: 4 BIRADS and only Low and High densities. For the two scenarios, we

combined all the images under the same cancer type and chose 100 images randomly. Here, the focus is on cancer detection without knowledge about the density type.

Experiment 4: Proposed two stage classification system

This experiment focuses on our proposed two stage approach which includes density classification and cancer detection. The two stages work in a sequential manner where the density classification is applied first, then cancer detection is performed. We start with 100 images chosen randomly from the dataset for each class. Then, the features obtained using features extraction techniques. Next, the features vectors are used for density classification to determine the density type. Based on density type, a dedicated classifier is used to identify the cancer type.

The experiments were developed under four scenarios. In the first scenario , we deal with low and high densities classes with all features (79 features). Then, we used the class assigned from the first stage which added to the feature vector to apply the second stage - cancer detection. In the cancer detection, we classified the images into either cancer exist or not. We used an optimal classifier for each class to achieve the best accuracy rate. Under the second scenario, we carried the same procedure as in the first scenario. Under the second scenario, we deal also with low and high densities but with top ranked features.

For the third and fourth scenarios, we deal with 4 BIRADS classes where the third scenario deals with all features and the fourth scenario deals with the top ranked features. The same procedure is followed as discussed above.

CHAPTER 4

RESULTS AND DISCUSSION

In this chapter, we discuss the results for density classification, cancer detection, and the combination of both using the proposed 2-stage classifier. We also compare our results with other state-of-the-art methods in the literature. Moreover, we highlight some challenges and difficulties we faced during the experiments.

Since the main contribution of our work is the 2-stage classifier which combines density classification and cancer detection, we prefer to start with discussions on our density classification experimental results. For the classification, we have used a 5-fold cross validation approach where we divide the dataset into 5 subsets, then use 4 subsets for training (80%) and 1 subset for testing (20%). This is repeated a number of times, then we average the results.

4.1 Density Classification

As mentioned earlier, there are two scenarios for breast density classification. The first scenario considers 4 BIRADs as follows:

- BIRAD I : the breast density is fatty.
- BIRAD II : the breast is fibroglandular dense.
- BIRAD III : the breast is heterogeneously dense.
- BIRAD IV : the breast is extremely dense.

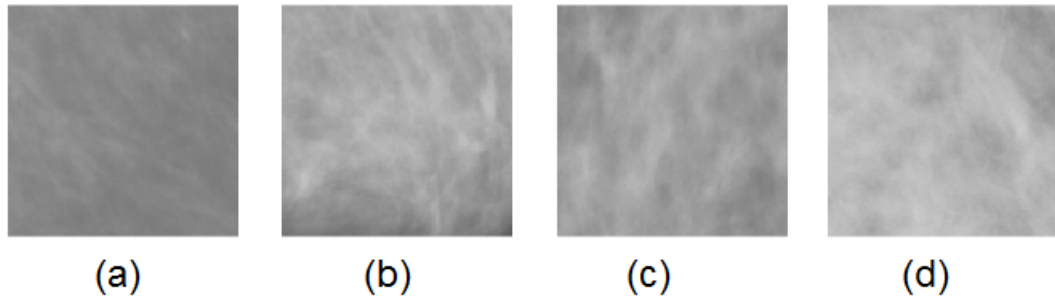


Figure 4.1: BIRADS examples: (a) BIRADS I, (b) BIRADS II, (c) BIRADS III, and (d) BIRADS IV

The second scenario only considers low and high density classes. For both types of breast density classification categories, we either use all extracted features directly or use the top ranked features. The effects of using both sets of features on classification performance are now discussed

★ Using all features:

Here, we use all the extracted features with four different classifiers that i.e. SMO, KNN, Naive Bayes, and Decision tree. The classification results are presented in Table 4.1.

Form Table 4.1, we observe that the SMO-based SVM gives the best performance among the other classifiers while the Naive Bayes did not perform well. The results are

Table 4.1: Density classification using all features

	SMO	kNN	Naive Bayes	Decision Trees
BIRADS (4 classes)	59.12%	52.47%	46.35%	51.9%
Low/High density (2 classes)	93.56%	92.7%	83.69%	91.63%

reasonable when we know that the assumption made on the independence of features in the Naive Bayes classifier is not realistic. Better results can be obtained when a general Bayesian classifier is used. Also, KNN and Decision Trees classifiers didn't give good performance as the KNN classifies the classes based on the distance between classes which not accurate since the features are not homogeneous in nature, and the decision tree classifier is mainly adopted for categorical data.

★ Using the top ranked features:

Instead of using all features for density classification, we have selected the top 20 high ranked features, using results from the Chi-Squared statistics and the Information Gain feature selection methods. We obtained different top ranked features for BIRADS breast density classes and Low/High breast density classes. The selected features for these two breast density categories are shown in Table 4.3 and 4.4, respectively. From these two tables, we note that some of the features are common across both breast density categories while due to the difference in textures in mammogram images for BIRADs classes, we have some different features selected as well (see Table 4.3 and 4.4). The classification results are shown in Table 4.2, where we notice that the SMO-based SVM again outperforms the other classifiers. The Naive Bayes also gives the same performance compared to the SMO-based SVM for Low/High breast density classification. The reason for this improvement in the Naive Bayes classifier is the selection of independent features which is a necessary requirement for the Naive Bayes

classifier. The Naive Bayes is mainly adapted to work with binary hypotheses as can be seen from these results. Recall that using feature selection using information content helps in selecting features that are relatively independent hence making the Naive Bayesian more applicable.

Table 4.2: Density classification using feature selection

	SMO	kNN	Naive Bayes	Decision Trees
4 BIRADS	55.36%	50.75%	46.14%	52.79%
Low and high density	92.06%	90.56%	92.06%	91.8%

Table 4.3: The top 20 features for 4 BIRADS classes

rank	feature approach	feature Name	IG	Chi
1	GLCM	Information measure of correlation 2	1	1
2	GLCM	Correlation	2	2
3	GLRL	RLN	3	3
4	GLCM	Sum entropy	4	5
5	image	Variance	5	4
6	GLCM	Information measure of correlation 1	6	8
7	GLCM	Cluster Shade	7	7
8	GLCM	Cluster Prominence	8	6
9	LBP	H(4)*	9	9
10	GLCM	Entropy	10	10
11	image	Skewness	11	11
12	LBP	H(9)	12	12
13	GLCM	Energy	13	13
14	GLCM	Maximum probability	14	14
15	GLCM	Difference entropy	15	15
16	GLCM	Inverse difference moment normalized	16	19
17	GLCM	Contrast	17	18
18	GLCM	Difference variance	18	17
19	LBP	H(10)	19	20
20	Wavelet	Energy of diagonal band in level 3	20	16

* H(x) = LBP Histogram bin number x

* IG: the ranking of features based on Information Gain, and Chi: the ranking of features based on Chi-Squared statistics.

Table 4.4: The top 20 features for low and high densities

rank	feature approach	feature Name	IG	Chi
1	GLCM	Information measure of correlation 2	1	1
2	GLCM	Correlation	2	2
3	GLCM	Information measure of correlation 1	3	4
4	GLCM	Cluster Shade	4	3
5	GLRL	RLN	5	6
6	image	Variance	6	5
7	GLCM	Cluster Prominence	7	8
8	GLCM	Sum entropy	8	7
9	image	Skewness	9	9
10	GLRL	SRN	10	10
11	GLCM	Sum of squares: Variance	11	12
12	GLCM	Autocorrelation	12	11
13	GLCM	Sum variance	13	14
14	GLRL	SGLGE	14	13
15	image	Mean of image	15	15
16	GLRL	LRE	16	16
17	GLCM	Sum average	17	17
18	GLRL	LRLGE	18	18
19	GLRL	SRHGE	19	19
20	LBP	LBP Histogram bin number 4	20	20

Summary:

The density classification plays an impotent role in cancer classification due to the strong relationship between breast density and cancer detection. We observed that the classification performance is different for breast density categories. For BIRADs classes, we used 128×128 patch size equals to 5mm x 5mm actual patch dimensions and we achieved 59.12% classification rate better than the results from [22] where they used full images for the feature extraction stage (see Table 4.5).

Table 4.5: Comparison for density classification with four BIRADS classes

Algorithm	classification rate
Oliver et.al [22]	47%
Proposed work	59.12%

For low and high densities classes, we compared our results with other methods in the literature as shown in Table 4.6. The results show that our proposed algorithm achieves an excellent classification accuracy of 93.6% using SVM, outperforming all exists approaches. Next, we will discuss the results related to cancer detection on

Table 4.6: Comparison for density classification with low and high densities classes

Algorithm	classification rate
Petroudi et al. [18]	91%
He et al. [20]	78%
Chen et al. [24]	88%
Oliver et al. [25]	91%
Ghouti and Owaidh [26]	83%
Proposed work	93.6%

mammogram images.

4.2 Cancer detection

For cancer detection, we classify the images under 2 different setups. In the first setup, we consider normal, benign and malignant mammogram patches, whereas in the second setup, we consider patches with and without cancer regions. We performed a number of experiments related to cancer detection. First, we mixed all types of densities and classified patches as either normal, benign or malignant patches using four different classifiers. The results are shown in Table 4.7. The maximum classification rate obtained was 55.54% using the SMO-based SVM with all features. The Naive Bayes gave the worst performance while the kNN and decision trees gave moderate performances. We also performed experiments to classify mammogram patches as either cancer or no cancer regions. We reached a maximum detection rate of 71.78%

using the SMO-based SVM. As expected, the classification accuracy has improved when considering only 2 classes.

Table 4.7: Classification rate for cancer detection for all images (BIRADS I,II,III,IV) (without density information , all features)

	SMO	kNN	Naive Bayes	Decision Trees
Normal, Benign and Malignant	55.54%	48.82%	41.3%	49.64%
Cancer/No cancer	71.78%	69.78%	70.39%	70.8%

We then considered the top ranked features in the next set of experiments. In particular, we used the top 16 ranked features shown in Table 4.10. In Table 4.8, the results were obtained for both Low and High density patches (BIRADS I and IV) separately under setup 1 (normal, benign and malignant patches). We reached a maximum detection rate of 57.8%, and 53.8% using the SMO-based SVM for Low density and High density, respectively. Overall, a maximum average classification accuracy of 55.8 % was reached when the type of density is known (only 2 types of density are considered). The results in Table 4.9 are similar to the ones in Table 4.8, however, only Cancer/No cancer classes are considered. We reached a maximum detection rate of 97% and 72.7% using the SMO-based SVM for Low density and High densities, respectively. Overall, a maximum average classification accuracy of 84.9 % was achieved.

Table 4.8: Classification rate for Cancer detection based on Low and High densities for setup 1 (Normal ,benign and malignant)(known densities, selected features).

	SMO	kNN	Naive Bayes	Decision Trees
Low density	57.8%	51.6%	50.8%	44.64%
High density	53.8%	47.8%	46.6%	45.2%
Low/High densities average	55.8%	49.7%	48.7%	44.9%

Table 4.9: Classification rate for Cancer detection based on Low and High densities for setup 2 (Cancer/No cancer)(known densities,selected features).

	SMO	kNN	Naive Bayes	Decision Trees
Low density	97.0%	95.0%	89.0%	95.0%
High density	72.7%	67.3%	68.0%	68.0%
Low/High density average	84.9%	81.2%	78.5%	81.5%

Table 4.10: The optimal features for cancer classification using High and Low density classes

rank	feature approach	feature Name
1	GLCM	Information measure of correlation 2
2	GLCM	Correlation
3	GLCM	Information measure of correlation 1
4	GLCM	Cluster Prominence
5	image	Variance
6	GLCM	Cluster Shade
7	GLCM	Sum entropy
8	GLRL	RLN
9	image	Skewness
10	LBP	LBP Histogram bin number 4
11	GLCM	Entropy
12	LBP	LBP Histogram bin number 2
13	LBP	LBP Histogram bin number 1
14	GLCM	Maximum probability
15	image	Kurtosis
16	LBP	LBP Histogram bin number 10

From the results in Tables 4.8 and 4.9, we observe that cancer detection performance is greatly effected by density classification. We achieved a maximum detection rate of 97% for low or fatty breast density patches and 72.7% for dense or high density patches. The average classification rate into cancer/no cancer patches is 84.9%. This result is much better than the case of unknown density type (Table 4.7).

The main reasons for the low classification accuracy in the case of high density are: the tumor size in mammogram patch, and the similarity of tumors in dense tissue. For the first reason, the full size of the patch is 128×128 with $50\mu/pixel$, so it is 5mm

x 5mm and the micro-classification size of the tumor is 0.5 - 1mm. Thus, the ROI area from our patch is only 1mm x 1mm which represents 1/25 from full patch size. The second reason is that the tumor looks like dense tissue or might be brighter and this problem complicates the diagnosis of tumors by radiologists leading to a wrong diagnosis or a miss. This problem is less prominent in the case of low density tissue.

4.3 The proposed two stage classifier

Since accuracy in cancer classification is substantially effected by the type of breast density, we propose here to combine breast density classification and cancer detection in a two-stage classification system. The experiments were performed for both the full set of features and the top ranked features. We list in Figure 4.2 the different experimental setups and show, as Experiment 4, how our approach is different to existing techniques. In particular, instead of ignoring completely the information on density or assuming it to be known, we propose to, first, identify the type of density we have, then, based on this stage, we use the specific classifier dedicated for this type of density.

Experimental results using all features:

Here, all features (79 features) are used for the experiments. For density classification (see experiment 1), we achieved a classification accuracy of 93.6 % for Low and High densities as shown earlier in Table 4.1. For cancer detection, we perform two experiments. In the first experiment, we performed cancer detection with prior- knowledge

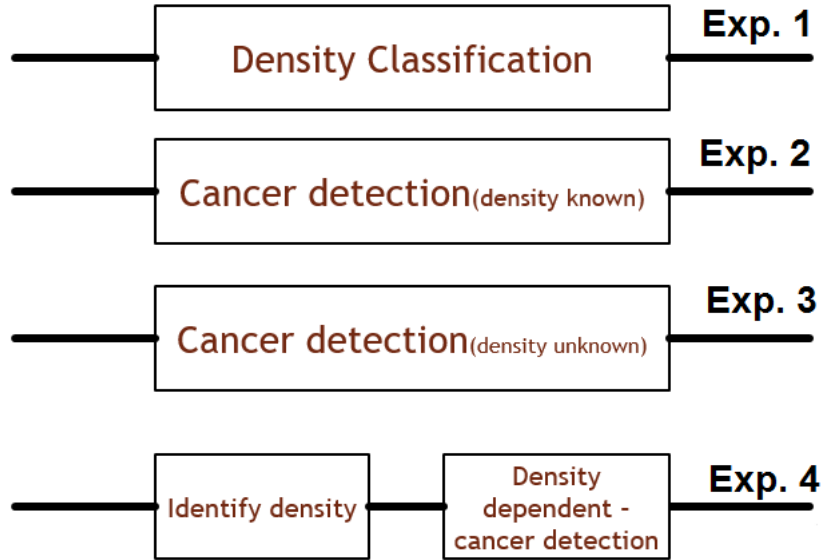


Figure 4.2: Experiments developed for our proposed algorithm

for density type (see experiment 2). We achieved classification rate of 83.9 % using the SMO-based SVM classifier for Cancer/No cancer classes. In the second experiment, we achieved a classification accuracy of 71.8 % using the SMO-based SVM classifier. In the experiment, no information about density is taken into the account (see experiment 3). The classification was also considered for Cancer/No cancer classes, too. All the above experiments have been performed with all features (79 features) . For experiment 4, the density classification was applied first, then we use the assigned classes in the next stage (cancer classification). Overall, for in the two stage classifier, we achieved an average classification rate of 78.5%

Experimental results using the top 16 ranked features:

Here, only the top ranked features (16 features) are used for the experiments. We performed the same experiments discussed above but with the top ranked 16 features.

For cancer detection, we performed cancer detection with prior- knowledge for density type (see experiment 2) and achieved a classification rate of 84.9 % for Cancer/No cancer classes as shown in Table 4.9. We performed the same second experiment as discussed above and achieved the same result. The experiments are also performed under Cancer/No cancer classes, too.

From experiment 4, the density classification was applied, we use the resulted classes in the cancer classification. For the case of the proposed two stage classifier, we achieved an average classification rate of 79.4%

The results are summarized for 2-stage classifier in Table 4.11. From the results, we observe an increase in classification performance using a small set of optimal features. The detection rate is increased by 0.9% using only the 16 top ranked features out of 79 features. This can greatly reduce the computational cost and improve the computational performance of our proposed approach. Also, we observe an increase in accuracy when prior knowledge about density classification is used (here, unknown but estimated) before cancer classification.

Table 4.11: Summery of 2-stage classifier

	Using all features	using feature selection
Cancer/No cancer classification without knowing density type (L/H densities, selected features)	71.8%	71.8%
Low/High density classification	93.6%	93.6%
Cancer detection with prior knowledge of density classes	83.9%	84.9%
Proposed 2-stage classification	78.5%	79.4%

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Breast cancer detection is still a challenging topic in medical image processing as breast cancer is a major threat to women's well-being. Numerous approaches and systems have been developed to assist doctors and radiologists in the early detection of cancer. CAD systems developed so far do not satisfy the clinical standards which deal with human lives. Current research activities focus on special problems such as classification between mass/no mass in ROI or between malignant and benign patches. Moreover, density classification is usually separated from cancer detection. In our proposed work, we combined density classification and cancer detection into a single robust CAD system. In our work, we started by proposing an approach for identifying the optimal set of features for cancer detection from small patches. Then, breast density and

cancer detection are fused into a single two stage classifier. The classification rate achieved around 80% without any prior knowledge about the density. We achieved an excellent result in density classification comparing to exists approaches reaching 93.58% accuracy with low and high density classes. Overall, we proposed a new framework for cancer detection which does not require any knowledge about the type of density in the mammogram images.

5.2 Future work

In cancer detection, many challenges still face researchers are faced the researcher. Following our research finding, These are some suggestions for future work:

- Building a robust algorithm for extracting ROI to improve density and cancer classification.
- Building a robust algorithm for extracting suspicious areas for abnormal masses in breast areas using MLO and CC images.
- Using segmentation methods over ROI to improve density and cancer classification.
- Building a robust approach to combine shape and textural features over ROI to identify malignant and benign masses.

REFERENCES

- [1] (2015, Nov.) Doral mri & radiology. [Online]. Available: <http://doralMRI.com/mamograms.html>
- [2] (2015, Nov.) Radiopaedia.org. [Online]. Available: <http://radiopaedia.org/cases/labelled-normal-mammograms>
- [3] G. Liasis, C. Pattichis, and S. Petroudi, “Combination of different texture features for mammographic breast density classification,” in *Bioinformatics & Bioengineering (BIBE), 2012 IEEE 12th International Conference on*. IEEE, 2012, pp. 732–737.
- [4] A. M. Sabu and D. N. Ponraj, “Poongodi. textural features based breast cancer detection: A survey,” *Journal of Emerging Trends in Computing and Information Sciences*, vol. 3, no. 9, pp. 1329–1334, 2012.
- [5] S. K. Biswas and D. P. Mukherjee, “Recognizing architectural distortion in mammogram: a multiscale texture modeling approach with gmm,” *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 7, pp. 2023–2030, 2011.

- [6] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.
- [7] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674–693, 1989.
- [8] I. Daubechies *et al.*, *Ten lectures on wavelets*. SIAM, 1992, vol. 61.
- [9] D. Raba, A. Oliver, J. Martí, M. Peracaula, and J. Espunya, “Breast segmentation with pectoral muscle suppression on digital mammograms,” in *Pattern Recognition and Image Analysis*. Springer, 2005, pp. 471–478.
- [10] (2015, Nov.) World health organization. [Online]. Available: <http://www.who.int/cancer/detection/breastcancer/en/>
- [11] (2015, Nov.) Saudi cancer registry. [Online]. Available: http://www.scr.org.sa/?module=publications&page=list&id=46&page_num=1
- [12] (2015, Nov.) American cancer society. [Online]. Available: <http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics>
- [13] (2015, Nov.) Susan g. komen. [Online]. Available: <http://ww5.komen.org/BreastCancer/Mammography.html>

- [14] R. L. Birdwell, D. M. Ikeda, K. F. OShaughnessy, and E. A. Sickles, “Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection 1,” *Radiology*, vol. 219, no. 1, pp. 192–202, 2001.
- [15] I. Kumar, J. Virmani, and H. Bhadauria, “A review of breast density classification methods,” in *Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on*, March 2015, pp. 1960–1967.
- [16] (2015, Nov.) The mini-mias database of mammograms. [Online]. Available: <http://peipa.essex.ac.uk/info/mias.html>
- [17] (2015, Nov.) Ddsm: Digital database for screening mammography. [Online]. Available: <http://marathon.csee.usf.edu/Mammography/Database.html>
- [18] S. Petroudi, T. Kadir, and M. Brady, “Automatic classification of mammographic parenchymal patterns: A statistical approach,” in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, vol. 1. IEEE, 2003, pp. 798–801.
- [19] A. Oliver, J. Freixenet, R. Martí, and R. Zwiggelaar, “A comparison of breast tissue classification techniques,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*. Springer, 2006, pp. 872–879.
- [20] W. He, E. Denton, and R. Zwiggelaar, “Mammographic segmentation and risk classification using a novel binary model based bayes classifier,” *Breast Imaging*, pp. 40–47, 2012.

- [21] M. Mustra, M. Grgic, and K. Delac, “Feature selection for automatic breast density classification,” in *ELMAR, 2010 PROCEEDINGS*. IEEE, 2010, pp. 9–16.
- [22] A. Oliver, J. Freixenet, and R. Zwigelaar, “Automatic classification of breast density,” in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2. IEEE, 2005, pp. II–1258.
- [23] S. Kutluk and B. Günsel, “Tissue density classification in mammographic images using local features,” in *Signal Processing and Communications Applications Conference (SIU), 2013 21st*. IEEE, 2013, pp. 1–4.
- [24] Z. Chen, E. Denton, and R. Zwigelaar, “Local feature based mammographic tissue pattern modelling and breast density classification,” in *Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on*, vol. 1. IEEE, 2011, pp. 351–355.
- [25] A. Oliver, X. Lladó, E. Pérez, J. Pont, E. R. Denton, J. Freixenet, and J. Martí, “A statistical approach for breast density segmentation,” *Journal of digital imaging*, vol. 23, no. 5, pp. 527–537, 2010.
- [26] L. Ghouti and A. H. Owaidh, “Nmf-density: Nmf-based breast density classifier.”
- [27] K. Ganesan, U. Acharya, C. Chua, L. Min, K. Abraham, and K. Ng, “Computer-aided breast cancer detection using mammograms: A review,” *Biomedical Engineering, IEEE Reviews in*, vol. 6, pp. 77–98, 2013.

- [28] Y. A. Reyad, M. A. Berbar, and M. Hussain, "Comparison of statistical, lbp, and multi-resolution analysis features for breast mass classification," *Journal of medical systems*, vol. 38, no. 9, pp. 1–15, 2014.
- [29] M. Asad, N. Z. Azeemi, M. F. Zafar, S. Naqvi *et al.*, "Early stage breast cancer detection through mammographic feature analysis," in *Bioinformatics and Biomedical Engineering, (iCBBE) 2011 5th International Conference on*. IEEE, 2011, pp. 1–4.
- [30] N. Gargouri, A. D. Masmoudi, D. S. Masmoudi, and R. Abid, "A new glld operator for mass detection in digital mammograms," *Journal of Biomedical Imaging*, vol. 2012, p. 4, 2012.
- [31] A. Oliver, X. Lladó, J. Freixenet, and J. Martí, "False positive reduction in mammographic mass detection using local binary patterns," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pp. 286–293, 2007.
- [32] S.-T. Luo and B.-W. Cheng, "Diagnosing breast masses in digital mammography using feature selection and ensemble methods," *Journal of medical systems*, vol. 36, no. 2, pp. 569–577, 2012.
- [33] J. NAGI, "The application of image processing and machine learning techniques for detection and classification of cancerous tissues in digital mammograms," Master's thesis, UNIVERSITY OF MALAYA, KUALA LUMPUR, 2011.
- [34] X. Xie and M. Mirmehdi, "A galaxy of texture features," *Handbook of texture analysis*, pp. 375–406, 2008.

- [35] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 6, pp. 610–621, 1973.
- [36] L.-K. Soh and C. Tsatsoulis, "Texture analysis of sar sea ice imagery using gray level co-occurrence matrices," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 37, no. 2, pp. 780–795, 1999.
- [37] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Canadian Journal of remote sensing*, vol. 28, no. 1, pp. 45–62, 2002.
- [38] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer graphics and image processing*, vol. 4, no. 2, pp. 172–179, 1975.
- [39] P.-Y. Lin, "An introduction to wavelet transform," *Graduate Institute of Communication Engineering National Taiwan University, Taipei, Taiwan, ROC*, 2007.
- [40] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *JOSA A*, vol. 2, no. 7, pp. 1160–1169, 1985.
- [41] L. Shen, L. Bai, and M. Fairhurst, "Gabor wavelets and general discriminant analysis for face identification and verification," *Image and Vision Computing*, vol. 25, no. 5, pp. 553–563, 2007.
- [42] M. Hanmandlu, A. A. Khan, and A. Saha, "A novel algorithm for pectoral muscle removal and auto-cropping of neoplastic area from mammograms," in *Compu-*

- tational Intelligence & Computing Research (ICCCIC), 2012 IEEE International Conference on.* IEEE, 2012, pp. 1–5.
- [43] N. Otsu, “A threshold selection method from gray-level histograms,” *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.
- [44] (2015, Nov.) Regionprops function in matlab. [Online]. Available: <http://www.mathworks.com/help/images/ref/regionprops.html?searchHighlight=regionprops>
- [45] (2015, Nov.) Matlab software. [Online]. Available: <http://www.mathworks.com/index.html>
- [46] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
- [47] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The weka data mining software: an update,” *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [48] G. Forman, “An extensive empirical study of feature selection metrics for text classification,” *The Journal of machine learning research*, vol. 3, pp. 1289–1305, 2003.
- [49] J. Platt, “Fast training of support vector machines using sequential minimal optimization,” in *Advances in Kernel Methods - Support Vector Learning*, B. Schoelkopf, C. Burges, and A. Smola, Eds. MIT Press, 1998. [Online]. Available: <http://research.microsoft.com/~jplatt/smo.html>

- [50] S. Keerthi, S. Shevade, C. Bhattacharyya, and K. Murthy, “Improvements to platt’s smo algorithm for svm classifier design,” *Neural Computation*, vol. 13, no. 3, pp. 637–649, 2001.
- [51] T. Hastie and R. Tibshirani, “Classification by pairwise coupling,” in *Advances in Neural Information Processing Systems*, M. I. Jordan, M. J. Kearns, and S. A. Solla, Eds., vol. 10. MIT Press, 1998.
- [52] D. Aha and D. Kibler, “Instance-based learning algorithms,” *Machine Learning*, vol. 6, pp. 37–66, 1991.
- [53] P. Cunningham and S. J. Delany, “k-nearest neighbour classifiers,” *Multiple Classifier Systems*, pp. 1–17, 2007.
- [54] R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers, 1993.
- [55] S. R. Safavian and D. Landgrebe, “A survey of decision tree classifier methodology,” *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [56] G. H. John and P. Langley, “Estimating continuous distributions in bayesian classifiers,” in *Eleventh Conference on Uncertainty in Artificial Intelligence*. San Mateo: Morgan Kaufmann, 1995, pp. 338–345.
- [57] I. Rish, “An empirical study of the naive bayes classifier,” in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22. IBM New York, 2001, pp. 41–46.

- [58] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok *et al.*, “The mammographic image analysis society digital mammogram database,” in *Excerpta Medica. International Congress Series*, vol. 1069, 1994, pp. 375–378.
- [59] T. M. Deserno, M. Soiron, J. E. de Oliveira, and A. d. A. Araújo, “Computer-aided diagnostics of screening mammography using content-based image retrieval,” in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2012, pp. 831 527–831 527.

Appendix

This is the feature vector that used in our proposed method.

SN	Feature approach	Feature Name
1	image	Mean of image
2	image	Variance of image
3	image	kurtosis of image
4	image	Skewness of image
5	GLCM	Autocorrelation of GLCM
6	GLCM	Contrast of GLCM
7	GLCM	Correlation of GLCM
8	GLCM	Cluster Prominence of GLCM
9	GLCM	Cluster Shade of GLCM
10	GLCM	Dissimilarity of GLCM
11	GLCM	Energy of GLCM
12	GLCM	Entropy of GLCM
13	GLCM	Homogeneity of GLCM
14	GLCM	Maximum probability of GLCM
15	GLCM	Sum of squares: Variance of GLCM
16	GLCM	Sum average of GLCM
17	GLCM	Sum variance of GLCM
18	GLCM	Sum entropy of GLCM

19	GLCM	Difference variance of GLCM
20	GLCM	Difference entropy of GLCM
21	GLCM	Information measure of correlation 1 of GLCM
22	GLCM	Informaiton measure of correlation 2 of GLCM
23	GLCM	Inverse difference normalized (INN) of GLCM
24	GLCM	Inverse difference moment normalized of GLCM
25	LBP	LBP Histogram bin (1)
26	LBP	LBP Histogram bin (2)
27	LBP	LBP Histogram bin (3)
28	LBP	LBP Histogram bin (4)
29	LBP	LBP Histogram bin (5)
30	LBP	LBP Histogram bin (6)
31	LBP	LBP Histogram bin (7)
32	LBP	LBP Histogram bin (8)
33	LBP	LBP Histogram bin (9)
34	LBP	LBP Histogram bin (10)
35	GLRL	SRE
36	GLRL	LRE
37	GLRL	GLN
38	GLRL	RLN
39	GLRL	RP

40	GLRL	LGRE
41	GLRL	HGRE
42	GLRL	SGLGE
43	GLRL	SRHGE
44	GLRL	LRLGE
45	GLRL	LRHGE
46	Wavelet Transform	Energy of image level 3
47	Wavelet Transform	Energy of horizontal band level 1
48	Wavelet Transform	Energy of horizontal band level 2
49	Wavelet Transform	Energy of horizontal band level 3
50	Wavelet Transform	Energy of Vertical band level 1
51	Wavelet Transform	Energy of Vertical band level 2
52	Wavelet Transform	Energy of Vertical band level 3
53	Wavelet Transform	Energy of Diagonal band level 1
54	Wavelet Transform	Energy of Diagonal band level 2
55	Wavelet Transform	Energy of Diagonal band level 3
56	Gabor filter	Energy of Band 1 from Gabor with 3 scales and 8 orientation
57	Gabor filter	Energy of Band 2 from Gabor with 3 scales and 8 orientation
58	Gabor filter	Energy of Band 3 from Gabor with 3 scales and 8 orientation
59	Gabor filter	Energy of Band 4 from Gabor with 3 scales and 8 orientation
60	Gabor filter	Energy of Band 5 from Gabor with 3 scales and 8 orientation

61	Gabor filter	Energy of Band 6 from Gabor with 3 scales and 8 orientation
62	Gabor filter	Energy of Band 7 from Gabor with 3 scales and 8 orientation
63	Gabor filter	Energy of Band 8 from Gabor with 3 scales and 8 orientation
64	Gabor filter	Energy of Band 9 from Gabor with 3 scales and 8 orientation
65	Gabor filter	Energy of Band 10 from Gabor with 3 scales and 8 orientation
66	Gabor filter	Energy of Band 11 from Gabor with 3 scales and 8 orientation
67	Gabor filter	Energy of Band 12 from Gabor with 3 scales and 8 orientation
68	Gabor filter	Energy of Band 13 from Gabor with 3 scales and 8 orientation
69	Gabor filter	Energy of Band 14 from Gabor with 3 scales and 8 orientation
70	Gabor filter	Energy of Band 15 from Gabor with 3 scales and 8 orientation
71	Gabor filter	Energy of Band 16 from Gabor with 3 scales and 8 orientation
72	Gabor filter	Energy of Band 17 from Gabor with 3 scales and 8 orientation
73	Gabor filter	Energy of Band 18 from Gabor with 3 scales and 8 orientation
74	Gabor filter	Energy of Band 19 from Gabor with 3 scales and 8 orientation
75	Gabor filter	Energy of Band 20 from Gabor with 3 scales and 8 orientation
76	Gabor filter	Energy of Band 21 from Gabor with 3 scales and 8 orientation
77	Gabor filter	Energy of Band 22 from Gabor with 3 scales and 8 orientation
78	Gabor filter	Energy of Band 23 from Gabor with 3 scales and 8 orientation
79	Gabor filter	Energy of Band 24 from Gabor with 3 scales and 8 orientation

Vitae

Name: Mahmoud Marwan Alhelou

Nationality: Palestinian

Date of Birth: 24-4-1989

Email: *mm.alhelou@gmail.com*

Mobile : *00966562477356*

Permanent Address: Jeddah, Saudi Arabia

Education:

King Fahd University of Petroleum and Minerals (KFUPM)

M.Sc. in Electrical Engineering CGPA: 3.028 December, 2015

Islamic University Of Gaza (Gaza Strip)

B.S. in Computer Engineering CGPA : 81.14% - very Good.

Experience (Training and Work) :

Courses Trainer : 2 months

ITIL Foundation , **Global Knowledge** Oct. 2015

DBA SQL server , **Saudi Electronic University** Sept. 2015

Program Trainer: 90 days

iExcel Gifted Program as Trainer in Electrical Engineering , **Saudi Aramco** Aug. 2015

Work Experience:

King Fahd University of Petroleum and Minerals (KFUPM) 2012 - 2015

Taught courses: Microprocessor, Microcontroller and Electrical circuit.

Workshops :

The Fifth Engineering and Technology Forum at the Islamic university of Gaza

The The first ICT Symposium 2011, Palestine Technical College - Deer AlBalah.

The The Attending Google Events , TEDxCairo and a lot of Scientism Conference .

Publication :

Hand Gesture System, the 2nd International Conference on Information and Communication Systems (ICICS 2011), May 22-24, 2011 Amman, Jordan.
