COMPUTER AIDED DIAGNOSTIC SYSTEM FOR CLASSIFICATION OF MAMMOGRAMS

A Dissertation submitted in fulfillment of the requirements for the Degree of

MASTER OF ENGINEERING

in

Electronic Instrumentation & Control Engineering

Submitted by

DHARMESH SINGH
801451005

Under the Guidance of

Dr. M. D. SINGH
Assistant Professor, EIED

2016

Electrical and Instrumentation Engineering Department
Thapar University, Patiala
(Declared as Deemed-to-be-University u/s 3 of the UGC Act., 1956)
Post Bag No. 32, Patiala – 147004
Punjab (India)

DECLARATION

I hereby certify that the work which is presented in dissertation entitled, "Computer Aided Diagnostic System for Classification of Mammograms", in partial fulfillment of the requirements for the award of the degree of Master of Engineering in Electronic Instrumentation & Control, submitted to Electrical & Instrumentation Engineering Department of Thapar University, Patiala is as authentic record of my own work carried under the supervision of Dr. M. D. Singh. It refers others researcher's work which is duly listed in the reference section. The matter contained in this dissertation has not been submitted, neither in part nor in full to any other degree to any other university or institute except as reported in the text and references.

Place: Patiala
Date: 07/07/16

Dharmesh Singh
Dharmesh Singh
801451005

It is certified that the above statement made by the student is correct to the best of my/our knowledge and belief.

Date: 7/7/16

Dr. M. D. Singh
Electrical and Instrumentation Engineering Department
Thapar University, Patiala - 147004

Countersigned by:

Dr. Ravinder Agarwal
Professor and Head
Electrical and Instrumentation Engineering Department
Thapar University
Patiala – 147004

Dr. S.S. Bhatia
Dean of Academic Affairs
Thapar University
Patiala –147004

Dedicated to
My parents

ACKNOWLEDGEMENT

Foremost of all, I wish to express my gratitude to **Dr. M. D. Singh**, Assistant Professor, Electrical and Instrumentation Engineering Department, Thapar University Patiala, who has been a tremendous mentor and teacher. This thesis would not have been possible without his constant encouragement and support. His constructive criticism and passion for perfection has had a profound impact in my life. I have always enjoyed his way of explaining things in a simple and elegant manner. I am truly very fortunate to have the opportunity to work with him. I found this guidance to be extremely valuable.

I am also thankful to our Head of Department, **Dr. R. Agarwal** as well as PG coordinator, **Mr. Nirbhowjap Singh**, Assistant professor, Electrical and Instrumentation Engineering Department. I would like to thank the entire faculty and staff of Electrical and Instrumentation Engineering Department and my friends who devoted their valuable time and help me in all possible ways towards successful completion of this work. I thank all those who have contributed directly or indirectly to this work.

Lastly, I would like to thank my parents for their years of unyielding love and encourage. They have always wanted the best for me and I admire their determination and sacrifice.

DHARMESH SINGH

801451005

TABLE OF CONTENTS

LIST OF FIGURES

\

LIST OF TABLES

LIST OF ABBREVIATIONS

Main symbols and notations used in this study are listed below. Sometimes a symbol may have alternate meaning, but in such a case; the context is sufficient to avoid confusion

BI-RADS	Breast imaging reporting and data system
CAD	Computer aided diagnosis
CC	Cranio-Caudal
CFS	Correlation based feature selection
DDSM	Digital database for screening mammography
DWT	Discreet wavelet transform
FOS	First order statistics
FFDM	Full field digital mammography
GLCM	Gray level co-occurrence matrix
GLRLM	Gray level run length matrix
GUI	Graphical user interface
GWT	Gabor wavelet filters
K-NN	K-nearest neighbor
LTEM	Laws' texture energy measures
MLO	Medio lateral oblique
MLP	Multilayer perceptron
ROI	Region of interest
SFM	Statistical feature matrix
SVM	Support vector machine

ABSTRACT

Breast cancer is the second cause of fatality among all cancers for women. Subsequently the reason of breast cancer remains anonymous, primary prevention becomes impossible. The most usual breast cancers types are mass (density) and micro-calcification. Early detection of these types of lesions is one of the important issues for breast cancer control. Currently, x-ray mammography is the particular most active, low-cost, and highly sensitive method for spotting small lesions. Thus, an adequate Computer Aided Diagnosis (CAD) classification system is designed for breast tumor for supporting immature radiologists in the diagnosis procedure. In this study, author proposed an efficient CAD system to categorize database into normal, benign and cancer breast tissue types. We have been extracted different type of squared shaped ROIs manually from the middle part of the breast. Statistical texture features are extracted from these ROIs. Finally, classification task is completed using Multilayer Perceptron (MLP) classifier, Support Vector machine (SVM) classifier and k-nearest neighbor (k-NN) classifier followed by feature selection technique. The outcomes of all the set of features are compared on the basis of their accuracy. Then, mammograms are categorized into normal, benign and cancer using the same procedure. The another proposed algorithm aims to developing an efficient image processing based Graphical User Interface (GUI) for detection of micro-calcification by wavelet transform in early stage. The result of first algorithm shows that accuracy reaches upto 90% when combination of optimal features used as input. We have successfully detected breast tumor using GUI model.

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Breast cancer is the major cause of death amongst women across the world. Breast tumor originates in the internal coating of milk lobules in the breast. It is basically abandoned progress of anomalous cells. With the help of finest treatment, disease free existence fluctuates from 98% to 10% in 10 years. Treatment consist of biopsy, medications (hormone treatment and chemotherapy), and radiation. There are many types of breast cancer, some of them are: inherent makeup, aggressiveness and with different steps (spread). A latest summary by National Cancer Registry Programs state that 28-35% breast cancer from all cancers are spread among females in major cities (Delhi, Mumbai, Chennai etc.) and it is increasing fast in huge figures [1]. "The American Cancer Society estimates that around 1,658,370 women in the US will be identified with breast tumor, and due to it about 589,430 females died till 2015" [2]. In the circumstance of breast cancer, primary prevention looks unbearable meanwhile the reason of this disease still a secret [3]. It is considered that most effective way to raise the chance of save from disease by diagnosis and treatment in early stages [4-6].

Mammography is the major screening tool which is carried out for discovery of breast cancer in early stage. And by the use of mammography at least 30% drop in breast cancer losses [7]. But some of the breast lesions such as micro-calcification, breast masses, shape distortion, and irregularity between breasts may not be detected by screening mammography because it is very difficult to interpret the morphological features [8]. Dense breast parenchyma is highly challenging job for sustaining sensitivity of mammography which depreciates both recognition and classification tasks [9]. Mammography cannot show that a defected area is cancerous, benign and normal alone. Breast biopsy methods have been removed cancerous tissues by earlier investigation. There is much type of breast irregularities which are visible in mammograms. Breast density is considered as a strong indicator for the growth of breast cancer. Masscalcifications and shape irregularities are the other lesions in breast. On the other hand, variations of breast tissue density are based on intensity values; hence these kinds of breast

density lesions classification can be considered as the problem of intensity and texture based descriptors. In several types of pictures, texture is considered as important attributes. Texture is basically in the form of information/data of the arrangement found in an image. Texture is a kind of visual features that does not influenced by color or intensity. So texture features have been extensively used in image processing. Various images can be distinguished by their textures without any extra information [10]. Finally, irregularity, and shape distortion are also very significant and deformities are hard to detect.

Radiologists make his last conclusion about lesion by second reading of the mammogram images. Breast biopsy technique was the technique for reading the mammograms previously. This method had shown the false-positive results. Due to false positive results unnecessary biopsy expected that only 15-30% cases are shown to be malignant [11]. Many of research work [11]-[13] have been presented that 10–30% of the noticeable cancers are unnoticed. Hence it is key necessity to increase the efficiency of screening databases by developing an automatic classification system. This system is very useful for radiologists to classification of doubtful areas. This system also drops the avoidable biopsies. It is concluded that, actual tumor remains unobserved by using false-negative results

Hence, the present work proposed a computer-aided diagnosis (CAD) classification system, which can be used for increment the performance to detect the irregularities in breast by minimize the unused lesions. We improved the accuracy assessment by providing computable investigation [14].

1.2 OBJECTIVE OF DISSERTATION

The objective of this dissertation is:

(i). Segregating regions of interest (ROIs) from the mammogram images.
(ii). Texture and intensity features are to be extract from each ROIs.
(iii). Check out the accuracy after correlation based feature selection (CFS) method is used.
(iv). Find the appropriate size of ROI which has best accuracy among all ROIs.

(v). Find out the best classifier which has best performance among k-Nearest Neighbor (k-NN), Multilayer Perceptron (MLP), and Support Vector Machine (SVM).

(vi). Detection of micro-calcification in mammograms using wavelet transforms.

1.3 ORGANIZATION OF THESIS

Here we show the different chapters which includes the content of thesis:

Chapter 1: Includes some of the evidences about cancer and breast cancer; brief introduction about breast tissue density and objective of the work.

Chapter 2: Includes the literature survey, which describes the various works related to this field.

Chapter 3: Includes a brief summary of mammograms and mammography.

Chapter 4: Includes the proposed algorithm for classification purpose by using intensity and texture features models and different type of classifiers. This chapter also contains another algorithm which detects the micro-calcification by using wavelet transform in early stage.

Chapter 5: Includes results and discussion of the different experiments.

Chapter 6: Describes the principal conclusion of this dissertation and also include future work.

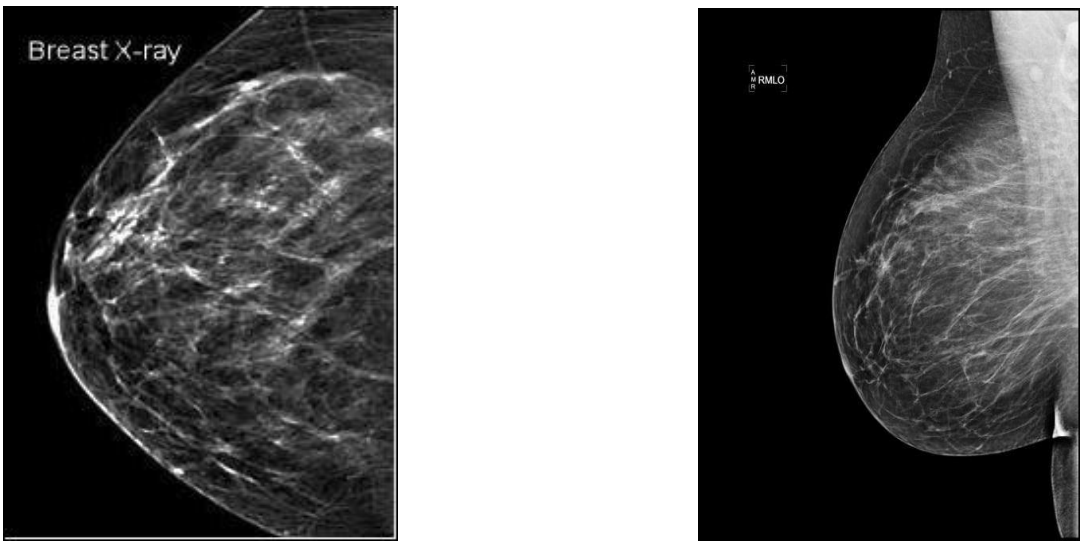
CHAPTER 2

MAMMOGRAMS AND DIGITAL MAMMOGRAPHY

2.1 MAMMOGRAM

X-ray exam of the breast is called a mammogram. That x-ray image i.e. mammogram is often used to detect breast tumor in early stage when tiny calcium cluster or any malignancy can be felt in it.

In late 1960's when proper x-ray machines were considered, modern mammography came into the picture. When we have used low dose of x-ray to identify the breast tumor in early stage, that type of breast imaging method is called Mammography. It is used before women sense symptoms- when it is most treatable. Nowadays mammograms are different from 1980s to 1990s due to technology has advanced a lot.



(a) (b)

Fig 2.1 (a) X-ray mammogram of breast and (b) Digital mammogram

In the present era, mammograms are achieving by low energy based x-ray machines. Routine chest x-ray or arm x-rays are not same as these low energy based x-rays, because these types of x-ray do not pass over tissues smoothly but quality of image improved. In today mammograms

expose too less radiation as compare from the past. Comparison of x-ray and digital mammogram has been shown in Fig 2.1 [15] [16].

2.2 TYPES OF MAMMOGRAMS

2.2.1 SCREENING MAMMOGRAMS

Screening mammograms are x-ray diagnosis of the breasts which are used for no breast lesions' women. The ultimate objective of a screening mammogram is to find breast tumor when it is in initial stage or it is too small to be perceived by a woman or her physician. With the help of screening mammogram, greatly improves the survival chance of recovery from breast cancer by finding the small breast tumors in early stage (before they have developed and spread).

Screening mammograms mostly associates two x-ray images (views) of each breast. Screening mammograms can also detect the microcalcification (tiny clusters of calcium) that also reason for the breast tumor.

2.2.2 DIAGNOSTIC MAMMOGRAMS

Diagnostic mammogram is used for those females, who have some breast lesions like nipple discharge or detect some malignant tissues. When some difficulty occurs in screening mammogram because existence of breast implants then diagnostic mammogram used to assess the variation formed during a screening mammogram. Diagnostics mammograms are executed for different goal but it is the same as normal x-ray exam. Other images are taken precisely research the area of interest during a diagnostic mammogram. In most of the cases certain images A diagnostic mammogram works as the reviewer for screening mammogram because diagnostic mammogram checks that area which is looked unusual on a screening mammogram is really defected or not. .

Diagnostic mammograms have more x-rays for views the picture of breast from different angles than screening mammogram, so these mammograms takes longer time. The specialist may enhance the defected area to accomplish a detailed image, which is very useful for the radiologist for make accurate diagnosis.

When some type of lesions occur in breast then radiologist advice to women for recheck normally in 4-6 months. It does not mean that individual has tumor, if the specialist mentions a biopsy surgery. Nearby 80% of all breast variations that are done by biopsy surgery, are originate to be benign (not cancer). There are many type of biopsy, patient should discuss with their specialist to choose which kind of biopsy is best. [17] [18].

2.3 MAMMOGRAPHY

Once a mammogram is finished, then part of breast is succinctly pressed among 2 plates closed to the mammogram apparatus. That type of apparatus consists of modifiable plastic plate (on top) and an immovable x-ray plate (on the lowermost). Occasionally the compression can sense troubled or painful and even hurting for specific females; it only takes little instants and is desired to acquire a decent picture. If any patients feel any type of pain, specialists reformed the subject to create the stress as easy as conceivable. The whole method for a mammogram takes near 20 minutes.

Dark and white x-ray image of breast muscle which is produced by instrument can be perceived on a processor display. These types of images totally depend on kind of instrument. These two methods of taking a mammogram are considering the equal. The instruments that acquire the mammogram on x-ray picture are screen filled units. For perceiving the picture on processor display we used Full-field digital mammography (FFDM) elements or units. Today FFDM units are used by many of mammogram machines. Fig 2.2 represents a full field digital unit.

3.1 HOW TO READ MAMMOGRAMS?

In mammography, interpretation of mammograms is relatively challenging task. The x-ray picture of the mammogram changes a countless deal from female to female. Besides certain breast malignancies may induce variations in the breast image that are tough to notice. If one of the female had past record of mammograms, it is very significant that the specialist has the newest x-ray pictures or digital pictures of her, as specialist can be matched with the new ones.

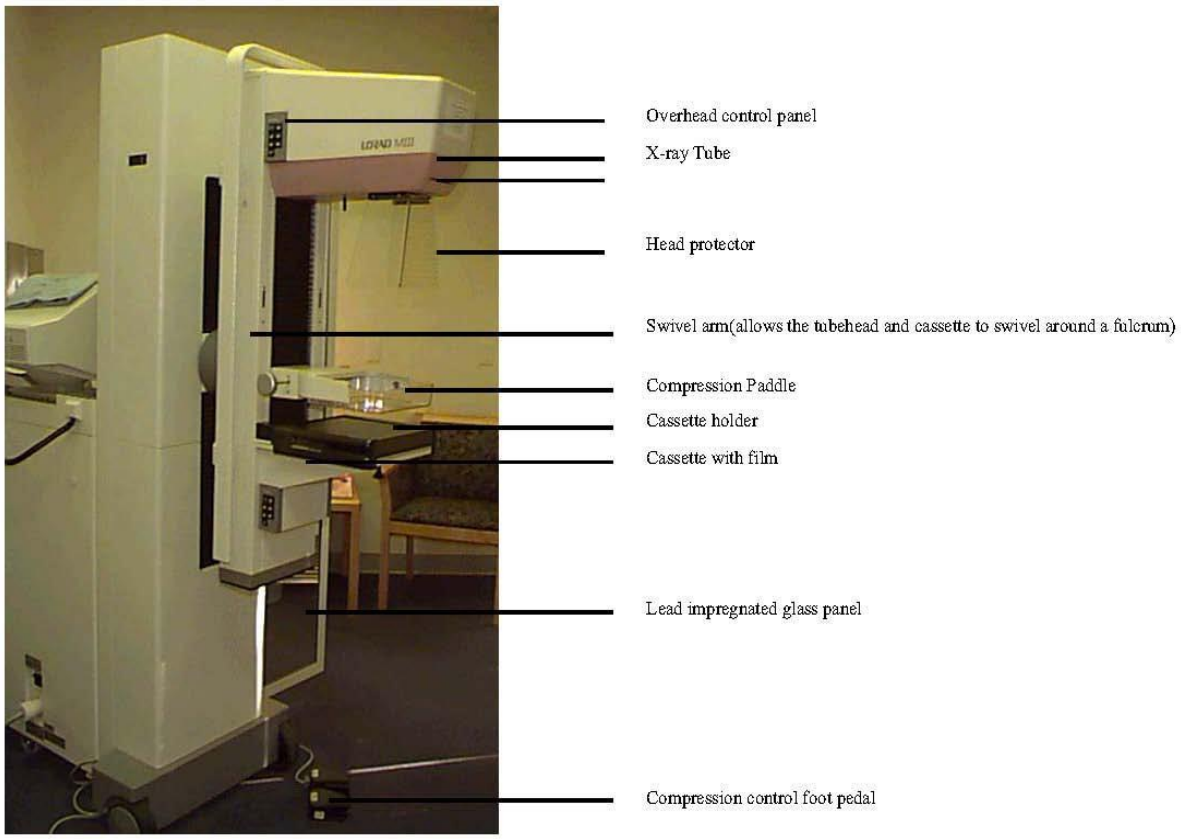


Fig. 2.2 Full Field Digital Mammography Unit Machine

By the comparing of new x-ray pictures to old pictures, doctor find minor variations that might have happened and discover tumor as early as possible. For better treatment, doctor suggested to patients to get mammograms with regular intervals of time.

2.4 MAMMOGRAM DEPICTIONS

2.4.1 BREAST DENSITY

The summary of the mammogram include of a valuation of breast thickness or state that individual has thick or dense breasts. Breast density depends upon three types of tissues, fatty tissue, fibrous tissue and glandular tissue.

If there is a thick or dense breast, they are not normal, as they are more chance towards the hazard of raising a cancer. While dense breast muscle can create it tougher to detect malignancies on a mammogram, at this interval, specialists do not decide what additional investigations, if any, must be completed in extension to mammogram high-risk collection. [19] [20].

2.4.1.1 BI-RADS FOR BREAST DENSITY

BI-RADS basically contain the assessment of breast density based on mammogram reports. BI-RADS categorize breast density into 4 groups:

BI-RADS 1: The breast is nearly completely fat and less than 25% fibrous or glandular tissues found in breast.

BI-RADS 2: These type of BI-RADS lesion distributed fibro glandular thicknesses. And it has 25-50% defected tissues.

BI-RADS 3: In this breast tissue is heterogeneously dense. It is tough to see small defected masses in the breast tissues. It has 51-75% malignant tissues.

BI-RADS 4: This means that breast have more than 75% fibrous and glandular tissue. If we are going to detect cancer in this type of category, it leads to losing some types of cancer cells. In certain conditions, the summary of the mammogram record is guided to patients, which contain info about breast thickness. The breast tissue is enormously dense in BI-RADS type 4.

That type of info may be called in unprofessional language in its place of the BIRADS types. BI-RADS types of breast density have 4 groups. Fig.2.3 represents the types of breast density basis of BI-RADS.

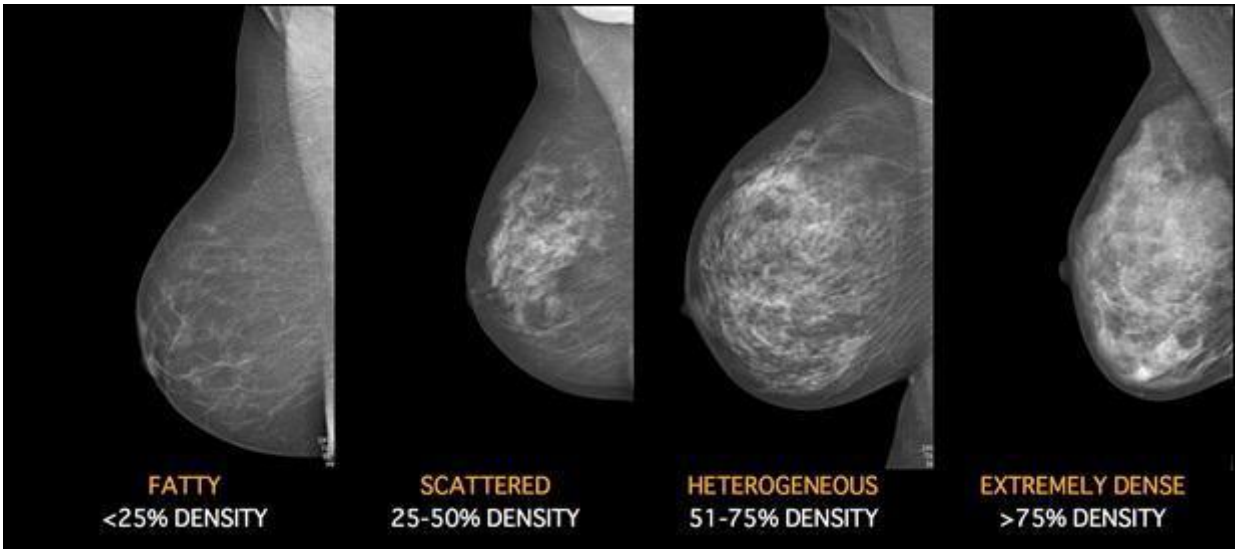


Fig. 2.3 Breast Density Category Basis of BI-RADS

2.4.2 CALCIFICATION

Calcifications are the tiny clusters of calcium. It is find in the milk lining or ducts in breast. These types of spot are highly intensity areas in mammograms. They might or might not be affected by cancer. There are 2 types of calcifications.

2.4.2.1 MACROCALCIFICATION

Due to aging of the breast arteries, swelling or old damages, some part of the breast is affected. In this part some uneven calcium deposits are found, that deposits are called macrocalcifications. These types of deposits do not need any surgery deposits as these are linked to non-cancerous circumstances.

2.4.2.2 MICROCALCIFICATION

Microcalcifications (shown in Fig 3.4) are little spots of calcium in the breast. These spots may be perceived alone or in masses (cluster). Micro-calcification may be abandoned and found fixed in a mass.

Individual micro-calcifications characteristically range in dimension from 0.1-1.0 mm with a normal diameter of around 0.5 mm. Micro-calcification are more concerned than macro-

calcification, because in macro-calcifications do not perpetually indicate that breast tumor is exist.. In most of the cases, if microcalcifications are present, it does not mean that biopsy is required. Radiologist estimates the probability of cancer from the contour and arrangement of the micro calcification. Question mark cancer is actually present in breast, biopsy is must for this type of situation [21].

.

.

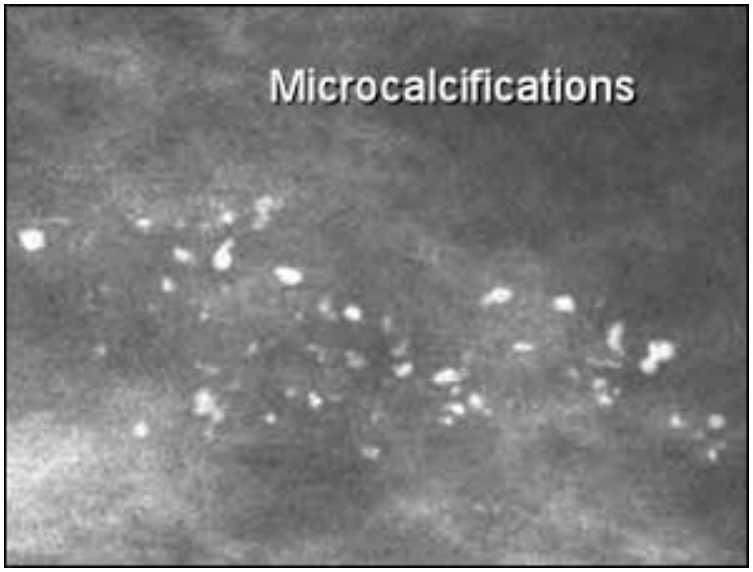


Fig 2.4 Microcalcifications in Mammograms.

2.4.3 MASS OR CYST

The parts which seen to be abnormal, are masses. Calcification is present in mass or not, it totally depends on reading of mammograms additionally. A mass consuming calcifications or not, is additional important modification reading on a mammogram. These types of part can be lots of depicts, including lumps or cysts (non-malignant, fluid-filled sacs) and non-malignant dense tumors (such as fibro adenomas). Figure 2.5 shows the different stage of mass in breast.

There are two types of cyst, first is fluid- filled cavity or sacs which is known as simple cysts and second is partly dense which is known as complex cysts

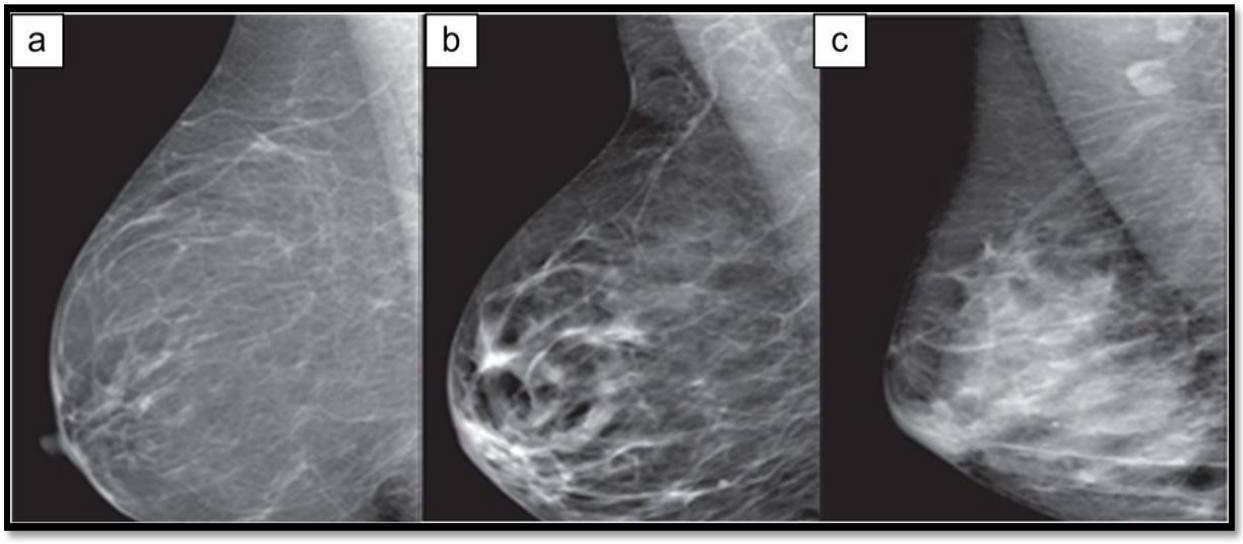


Fig. 2.5 (a) Normal Mass, (b) Benign Mass and (c) Cancer or Malignant Mass

Benign and normal mass type tissues are the simple cysts (harmless) and biopsied is not required. Malignant or cancer mass type tissues are complex cysts and might be requiring biopsy so that it does not become higher risk of breast cancer in future.

In physical exam, a cyst and a tumor are same in terms of density. Both of them also seem to be similar in breast x-ray. Breast ultrasound is often useful for the cysts detection and also confirms that mass is actually a cyst. Alternative choice is to pull out the liquid starting from the tumor with small and deep needle. And if mass are partly solid or complex cysts not simple cysts, additional imaging exams may be completed. Some of the masses can be treated by watching the normal mammograms, while another may want a biopsy.

There is some important parameter of tumor which is so needful for radiologists. Those are dimensions, shape, and boundaries (edges) of tumor. It is very important to know the initial stage of mammograms for better diagnosis. If mass or cysts not change in many years, it can be the cause of tumor. Figure 2.6 showing mammogram with a cyst.

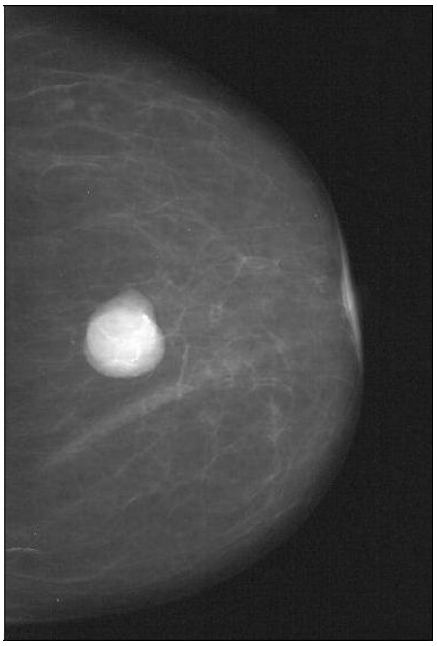


Fig 2.6 Mammogram with a cyst

2.5 LIMITATIONS OF MAMMOGRAMS

There are many limitations in mammography while we do many of medical tests. Nowadays screening of the breast cancer is the finest method to detect cancer in early stage. But detecting cancer in early stage does not always decrease the chance of breast cancer and considering a small growth of malignant tissues do not always say that it can be treated.

2.5.1 FALSE-NEGATIVE RESULTS

A false-negative mammogram looks ordinary even when breast cancer is exists. High breast thickness is the leading cause behind false-negative results. In all screening mammograms tests one is missing out in every five breast cancers. Younger women have dense breast compare to older women, thus false negative results occur more in younger women. Due to false-negative results treatment can be delayed.

.

2.5.2 FALSE-POSITIVE RESULTS

A false-positive breast x-ray images looks anomalous but in this no tumor is really exist. Anomalous mammograms need additional analysis (diagnostic mammograms, sometimes biopsy and ultrasound) to detect if cancer is existing. With yearly screening test, the probabilities that a lady will have false-positive results are more than 50% over a 10-year period. The probabilities of false-positive results are highest in starting mammogram, and are slightly decreasing on succeeding mammograms. False-positive breast x-ray exams can cause of impermanent nervousness. Still, many researches have exposed that "women receive false positive outcomes as portion of the method to detect breast cancer in early stage" [22].

.2.6 IMPROVING MAMMOGRAMS

Mammogram is an outstanding technique to detect most of the breast cancers in early stage, when growth of tumor is small and most treatable. But mammogram does not spot all breast cancers. New systems like computer aided detection are to make mammograms more precise.

2.6.1 COMPUTER-AIDED DETECTION AND DIAGNOSIS

To detect doubtful variations on mammograms more accurately, radiologists were established Computer-aided detection and diagnosis (CAD) system . Standard film mammograms and digital mammograms are used in this emerging technology.

By the CAD system doctors detect anomalous regions on a mammogram by using as another set of perception. Many of the early research on CAD system showed that it is very useful for the clear improvement for detecting small cancers. But studies related of the CAD system shows mixed results of the community practice. Some of them presented a strong advantage from the use of CAD system, and other studies presented that it did not detect more cancers or invent cancers in early stage. It is concluded that CAD classification system is only supportive when the radiologists are expert and also have proficiency in reading mammograms [23].

2.6.2 TOMOSYNTHESIS (3D MAMMOGRAPHY)

3-D mammography technology is basically addition of 3-dimensional pictures. In this test, the breast is squeezed once and an instrument receipts various low-dose x-rays as it transfers completed in the breast. A breast tomosynthesis instrument shown in Figure 2.7 was permitted in USA by the Food and Drug Administration (FDA) in 2011. Figure2.8 represents the contrast between a digital breast tomosynthesis and a traditional x-ray mammogram. It is clearly seen that the tomosynthesis digital mammogram is significantly clear than the traditional mammogram and represents the malignant tissue with considerable depth and transparency.

Fig 2.7 Tomosynthesis Machine

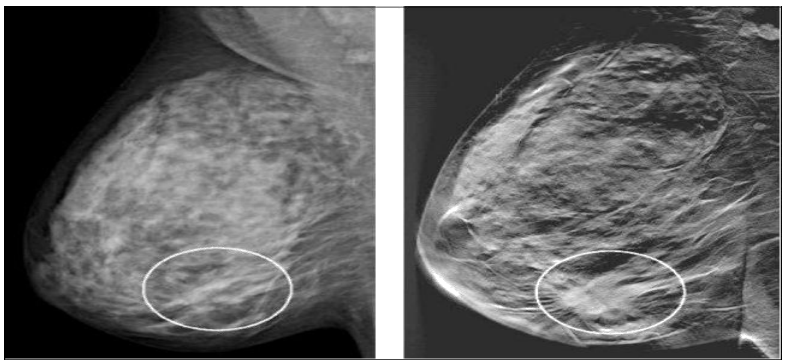


Fig 2.8 (a) Tradition x-ray mammogram and (b) Digital breast tomosynthesis

CHAPTER 3

LITERATURE REVIEW

This chapter includes discussion of the work done by numerous scholars in the area of Digital Mammography. They give several of their publications and involvement related to mammography. The main aim of all of the studies is to improve accuracy of the diagnosis which is useful for the radiologists to detect the breast tumor efficiently. Many of researchers were done work in the field of classification of mammograms on the basis of mass and breast density in past. These researches include the ROIs extraction, texture feature extraction, and many of machine learning classifiers used for classification. All of the texture features are worked as input vector in classifiers, and classified the mammogram database into three categories normal, benign and cancer.

3.1 REVIEWS FOR PERFROMANCE OF CLASSIFICATION (BASED ON BRAEST DENSITY)

Many of previous studies have presented that the performance accuracy for identifying a breast cancer drops due to development of breast thickness [24, 25]. Radiologists evaluate breast density by mammography screening tool which is highly useful for patient. Computer based classification system is needful for detection of breast lesions in dense mammograms. This computer based diagnosis system includes the various texture models. There are many studies which includes the various texture features which are used for classification.

Li et al. [26] have considered the size and location of region of interest (ROI), and also included the effect of size and location. They determined that the location of ROI plays important role for performance of texture features. The accuracy of performance decreases if the location of ROI is moved from the midpoint of breast area to the back part of nipple. Because that back part of breast includes the thick or dense region.

Bovis et al. [27] done the classification of mammogram task based on mass and density. The authors used 377 mammograms from DDSM database for their work. They used only four texture models for classification purpose. There are two unlike algorithms which adopt by the authors. First includes the two class problem and second includes the four class problem. ANN classifier used for the classification purpose. They achieved 71.4% accuracy.

Petroudi et al. [28] includes a method for involuntary classification of breast mass or density arrangements. These breast density arrangements inside the breast are shown by statistical features. In their technique, every mammogram is classified into three unlike parameters: breast tissue, background and pectoral muscle. In conclusion, an overall of 132 mammograms selected from the Oxford Database.

Hapfelmeier et al. [29] evaluated two computer aided diagnosis prototypes which includes segmentation, texture feature extraction, and classification for mass lesions and micro calcification for evaluating the hazard of breast cancer. The result includes classification performance of CAD prototype shown AUC=0.777 for 242 texture features. These analyses included 1347 ROIs on DDSM database. They used Support Vector Machine (SVM) for classification task and Linear Discriminant Analysis (LDA) used for feature selection task.

Oliver et al. [30] suggested a CAD method for classification of breast thickness using morphological and roughness features. In this two database sets used for evaluating the accuracy of proposed system. First mini-MIAS database includes 322 images and DDSM database includes 831 images. Bayesian classifier, Decision Tree classifier, and K-Nearest Neighbor were used for classification task.

There are many of texture features which basically depend upon the neighboring pixels, as Gray Level Co-Occurrence Matrix (GLCM) and Gray Level Run-Length Matrix (GLRLM), Haralick, [47] presented GLCM method in 1973. Many of investigators used the GLCM matrix for extracting the features which are useful for classification. 100 features extracted by Zarchari *et al.* [31] where they used contour based methods, statistical, GLCM, intensity, and Gabor.

Georgiadis *et al* [32] extracted 4 different features from histograms, 4 different and 10 features from GLRLM matrix and 22 features from GLCM matrix,.

Ciatto et al. [33] assessed the breast density by QUANTRA software in contrast with graphic classification. Initially, the authors examined the calculation of graphic thickness by eleven knowledgeable radiologists that categorize set of 418 mammograms. By QUANTRA software compare the two methods and deliver standards for regulating arithmetical computerized breast density valuation to presently used BI-RADS standard.

3.2 REVIEWS FOR DETECTION OF MICRO CLACIFICATION

There are many different methods that have been useful for the recognition of microcalcifications. The Wavelet transform is an ideal tool for detection of micro-calcification, as related to the Fourier transform. Because Fourier transform is deal with frequency content only. Therefore, wavelet transforms have used by several scholars for the recognition of microcalcifications [34]-[39]. A summary of certain approaches, which have been used the microcalcification, is specified below.

Strickland and Hahn [34] suggested a technique using biorthogonal wavelet transforms and sub-band allowance to spot grouped microcalcifications. Zhang *et al.* projected a procedure [35] to improve the masses at distinct methods of the wavelet disintegration. Qian *et al.* [36] included a non-linear filter for removing noise of image. That nonlinear filer is organized by tree based wavelet transform. It is used for multiresolution disintegration.

However, Gurcan et al. [37] included the method which produce high pass and low pas components from image. They try to solve the 3-dimensional sub band image area. Hence, the changes in the distribution of band pass and high pass decomposition filters based coefficients were computed by histogram based third order and fourth order moment. Recently, arithmeticians have proved that “wavelets have fixed square supports and are perfect for taking point incoherence, but not boundaries” [38]. It is concluded that wavelet based method for is successful for detection of microcalcification.

Other multi-scale methods have been inspected in addition to wavelets. Netsch and Peitgen [39] proposed a computer aided model which is includes Laplacian of Gaussian filter and a mathematical model. The authors used the multiscale detection methods. These methods try to detect the high intensity defected areas. They used Laplacian filtering for the detection of tiny defected clustered microcalcifications. The drawback of these methods is that if dense background is present then these approaches are more expected to fail.

Jasmine, J.S.L used a new method to detect the micro calcification using the combination of wavelet analysis of the image. And for the classification task artificial neural network is used. Micro-calcification is the high frequency component and tiny clusters of calcium. There is one input, two hidden layers and one output in ANN. In this study, the system is classified in two classes (normal and abnormal). The results of the experiment employ that detection rate reaches upto 87% and no false detection. It is very significant result. Here MIAS dataset has been used.

Subhash J. et al. [40] was presented a new technique for detection and categorization of micro calcifications. In this, Fuzzy c-means clustering (FCM) is applied for segmentation and two dimensional discrete wavelet transforms (DWT) are mined from the finding of micro calcifications. 9 statistical features are extracted from different ROIs and these input fed into Artificial Neural network (ANN). This process carried out in 322 images from MIAS database

Table-2.1 Summary of previous breast density classification in chronological order

Author	Image classes	No. of Images	Features	Classifiers	Accuracy (%)	Drawback
Karssemeijer et al. [41]	BIRADS-IV	615	First Order Statistics	k-NN	67	For the classification number of cases differed by another classes was only 2%
Bovis, et al. [27]	BIRADS-IV	377	SGLD Features, Fourier transform, LTEM and DWT textures	ANN	71.40	The four classes increase the probability of misperception in the subsequent classification.
Petroudi, et al. [28]	BIRADS-II	132	Statistical Distribution & gray level based features	K-mean algorithm	76	Invariably filter response is low due to arithmetical distribution of dimensional space.
Li et. al [26]	BRCA1/ BRCA 2	90	First order statistics texture features, Fourier and fractal analysis	Bayesian artificial neural network	72	Accuracy of classification significantly decreased with decreasing the ROI size
Oliver, et al.[30]	BIRADS 4-Class	831	Relative areas, center of mass, Intensity of both cluster	KNN, Decision tree, Bayesian classifier	77	Biasing occurred in methodology algorithms
Sampaio et al. [45]	BIRADS-II	623	Shape and texture features	Cellular neural network	80	Parameters are hard to understand. Complex parameters
Liu, et al.[44]	BIRADS-IV	88	FoS texture features	SVM	86.40	Numerous complex constraints in the proposed method prejudiced the final results.
Mustra et al [42]	BIRADS-IV	322	GLCM features	KNN(k=1)	79.30	Overfitting problem occurred due to complexity of model
Sharma, et al. [43]	Fatty and dense	322	SGLCM, GLDS, FoS, SFM, LTEM	SVM	89.00	Only for 2-class problem not more
Present work	BIRADS-III	480	Texture and Gabor wavelet features	SVM, MLP K-NN	90	----

3.2 SUMMARY OF PREVIOUS WORK (In Chronological Order)

The earlier work concludes that the classification of mammograms is based upon breast density mostly. Li *et al.* [26] discussed that location and dimension of ROIs play important role for classification as it include back part of nipple. But accuracy of classification decreased with decreasing the ROI size.

Karsseneijr *et al.* [41] found 67% accuracy for BIRADS-IV classification. K-NN classifier used for classification purpose. But classification of number of cases differed by another classes was only 2%. Bovis *et al.* [27] considered 4-classes and many of texture and fractal features, but in this probability of misperception in the subsequent classification is increased. Petroudi *et al.* [28] had shown 2-class problem and considered 132 mammogram images with 76% accuracy. But it has very low classification response using invariant filter.

Most studies have focused on classification of breast tissue based on mass and density. Hapfelmeier *et al.* [29] found significant changes in CAD system and 1347 ROIs are used it includes two CAD prototypes, texture feature extraction, breast lesions classification and micro calcification detection. It has shown 77.77% accuracy. This is comparatively low. Oliver *et al.* [30] achieved 77% accuracy with K-NN classifier. But biasing problem occurred in methodology algorithms. Sharma *et al.* [43] concluded the SVM is the best classifier for the classification of breast density and it had an accuracy of 89% when feature selection is used. But it was only for 2-class problem not more. In this thesis we considered different texture models, features selection algorithm and hybrid classifiers to develop the performance of classification of breast tissues as noted from literature review discussed above.

In present work, a CAD system is established for classification purpose of 3-class mammogram lesions problem, and evaluate the performance by accuracy among normal, benign and cancer classes. Here we have been used different texture models for intensity and texture features extraction. From the literature review, it was found that rare studies have been done for feature selection to enhance the accuracy. That's by we has been used correlation based feature selection (CFS) [46] to choose the optimal features from complex dataset by using best first search algorithm. For the classification purpose k-Nearest Neighbor (k-NN), Multilayer Perceptron (MLP), and Support Vector Machine (SVM) classifiers have been used. The classification accuracy reaches up to 90%. In present work, we also create a GUI model for detection of micro-calcification using wavelet transform in early stage

CHAPTER 4

MATERIAL & METHEDOLOGY

4.1 DATABASE

Digital Database for Screening Mammography (DDSM) [49] database is used for this study. The DDSM is openly available from the University of South Florida. We have been used 480 cases for this study. There are different types of file in DDSM database, first is "ics" file, second is four image files which are using lossless JPEG encoding, third is zero to four overlay files and last one is "16 bit PGM" files. There are two most communal methods of breast projection are Cranio-Caudal (CC) and Medio Lateral Oblique (MLO). The CC view is taken from above, so area close to the chest wall does not display. In MLO projection, x-ray is taken from front side of the breast as whole breast is visible. In the present work, overall 480 (160×3) mammograms (MLO views) including of 160 mammograms from every of the 3 categories of breast lesions diagnosis taken from the DDSM database. All images have 43.5 microns sampling rate and 16 bit gray levels.

4.2 PROJECTED METHODOLOGY (ALGORITHM 1)

We projected a Computer Aided Diagnostic (CAD) classification system (shown in Fig 4.1) which is used to enhance the performance of three lesions type classification. This system mainly discriminate the three lesions by using texture models. The importance of this CAD system is to solve out the all limitations which found in literature survey. There are four modules in CAD system, which are,

i. Region of interest (ROI) extraction module.
ii. Intensity and texture extraction module.
iii. Feature selection technique module for dimension reduction
iv. MLP, K-NN and SVM classifiers module

4.2.1 REGION OF INTEREST (ROI) EXTRACTION MODULE

We extracted the different size of ROIs manually from DDSM databases by developing a MATLAB algorithm. ROI(s) plays an important role in medical imaging. ROI contains crucial information associated to the analysis as a sub part of image. In case of tissues or organs, squared shaped window size ROI is more beneficial than the entire image [49]. By use of ROI, computational time will reduce. The following calculations and investigation extremely depend upon ROIs, so size and location of the ROIs very deciding. If we have large ROI size like 512×512 or small ROI size like 64×64, does not accommodate too much or too less information for calculation respectively[50]. Hence it is concluded that ideal size of the ROI confides in upon the proposed algorithm.

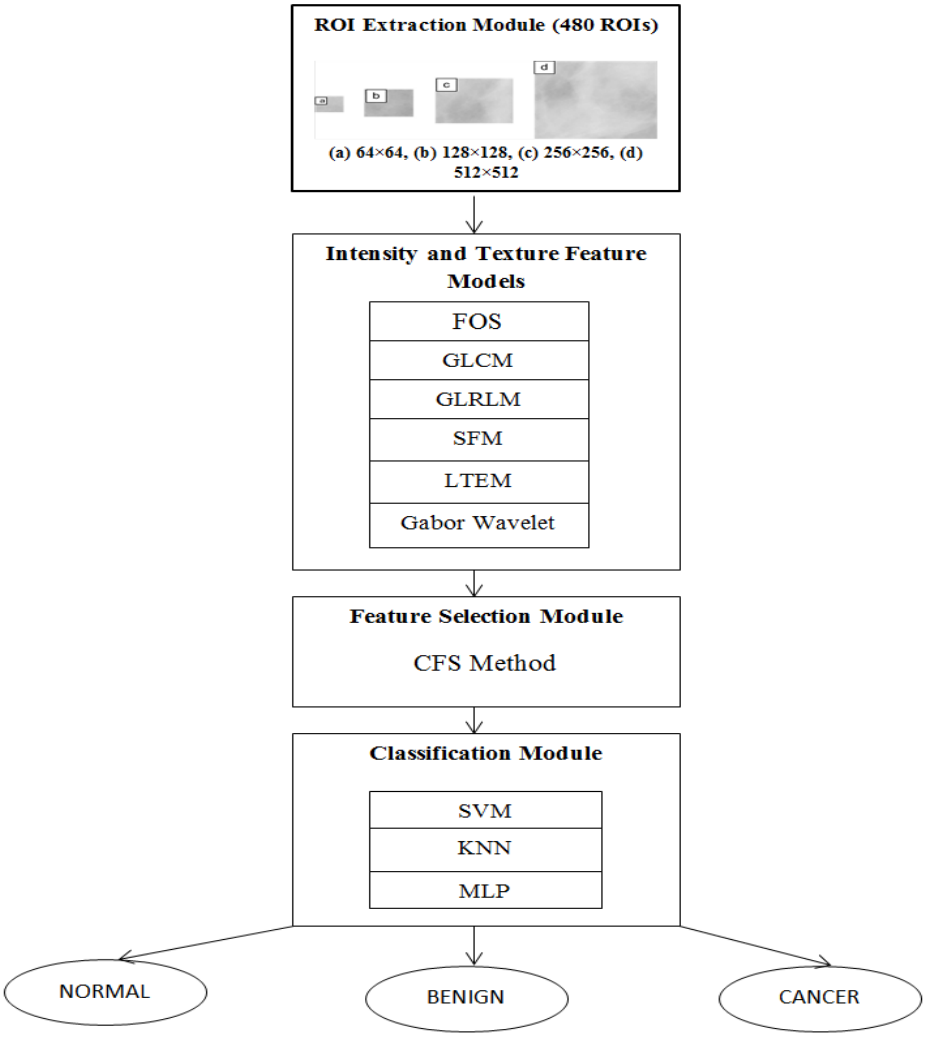


Fig. 4.1 Computer aided diagnosis algorithm for classification of 3-class problem

In the first step of the proposed CAD system algorithm, we have extracted different size of ROIs (64×64, 128×128, 256×256, and 512×512) from each class of mammogram. The ROIs extracted from the area right behind of nipple (as shown in Figure 4.2). Texture and intensity based extracted feature analysis generally evaluate the characteristics of regions of interest (ROIs). Identification of specific types of ROIs is easy, when mammograms are analyzed through center. In digital mammography, high resolution gray level images are obtained, where texture plays significant role. For identification purpose features extraction descriptors are desired, which is useful for characterize a mammogram. Really this is the good way to accomplish texture information with primary use to classify the mammogram data. The grey scale image defined by spatial distribution of the intensity pixels, which contain vital information for classification. Hence texture describes the repeating pattern of local variations of grey tones in an image.

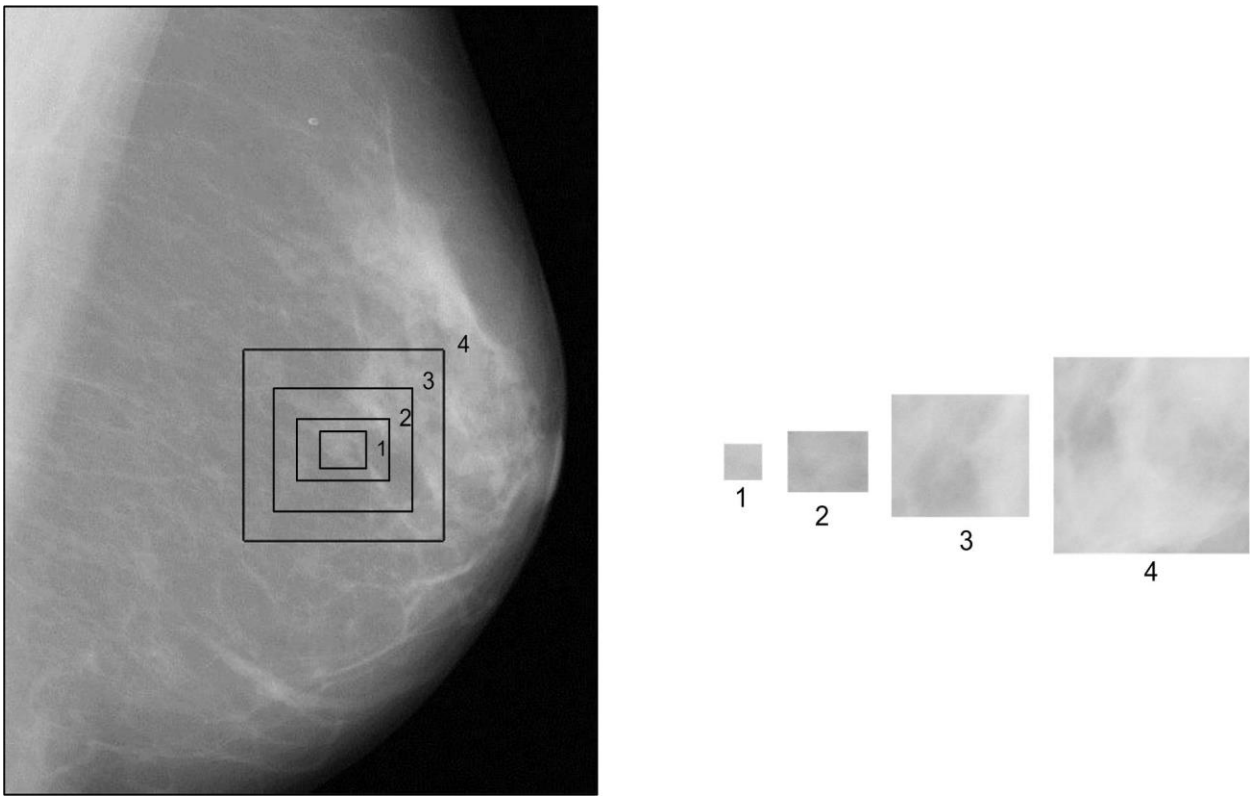


Fig. 4.2 Various Size of ROI Extraction (64×64, 128×128, 256×256, and 512×512)

4.2.2 INTENSITY & TEXTURE FEATURE MODELS

Features are defined by the combination of gray-level intensities. These intensities are carried out through every point of the image. The features who define points of an image depend upon the

features of categorization. There are different kind of features categorization, first order, second order and higher order features, where higher order defining by the higher number of features.

In the algorithm 1, for intensity and features extraction module, ROIs are act as input. Initially, six different intensity and texture features are extracted. Law's texture energy mask for spatial filtering and Gabor wavelet are selected for multi-scaled and multi-resolution tunable analysis. These six methods for features extraction are (i) Statistical Features Matrix (SFM) (ii) Gray Level Run-Length Matrix (GLRLM) (iii) Gabor Wavelet (GWT) (iv) Laws Texture Energy Measures (LTEM) and (v) First Order Statistics (FOS) (vi) Gray-Level Co-Occurrence Matrix (GLCM). There are many researchers [51, 52], who has been used combination of these features. On the base of these studies and information features selection is used. In the proposed CAD system, total of 361 perceptual and non-perceptual texture and intensity attributes are extracted. From six texture models total six feature sets are obtained. These feature sets were used as single feature vector and concatenated feature vector. For the next step, these feature vectors are used for the selection and classification of the image. These six textural feature models are:

4.2.2.1 FIRST ORDER STATISTICS

First order statistical features are determined from the spatial distribution of grey levels values of an image. FOS deals with the individual pixel of an image, not depends upon neighbourhood pixels. Hence histogram based first order statistics features are used in texture analysis. There are six features which are extracted from histogram based first order statistics. Let I(x y) is an image, and *P (I)* is defined as the first order histogram

$$P(I) = \frac{\text{Number of pixels with gray level I'}}{\text{Total number of pixels in an image}}$$

Table 4.1 represents the first order statistics based texture features and corresponding mathematical formulae:

Table 4.1 FOS features and their mathematical formulae

S.No.	FOS Features	Mathematical Formulae
1	Mean	$f_1 = \sum_{i=0}^{N_g-1} i.P(I)$
2	Variance	$f_2 = \sum_{i}^{N} (g - f_1)^2 P(I)$
3	Skewness	$f_3 = \frac{1}{f_2^3} \sum_{i}^{N} (i - f_1)^3 P(I)$
4	Kurtosis	$f_4 = \frac{1}{f_2^4} \sum_{i}^{N} (i - f_1)^3 P(I) - 3$
5	Energy	$f_5 = \sum_{i} P(I)^2$
6	Entropy	$f_6 = -\sum_{i} P(I) \ln P(I)$

4.2.2.2 GRAY LEVEL CO-OCCURANCE MATRIX:

First order statistics based texture features deal with the gray level distribution of an image. Nevertheless first order statistics based features do not tell about the information of gray level pair of pixel and relative location. The second order statistics based features based on the relation between to neighboring pixels in different offset or different angles, where first pixel is called as reference and other is called as neighbor pixel. Gray level co-occurrence matrix (GLCM) is used to compute the texture features based on second order statistics. [53]

In GLCM matrix amount of rows and columns is similar to amount of gray levels. The GLCM matrix of frequencies $P_{\theta,d}$ (I_1, I_2), show the repetition or occurrence of gray level pairs. In GLCM matrix, second order statistical probability values depend upon displacement distance '*d*' and particular angle 'θ' with gray level values of I_1 and I_2. The second order statistical attributes are selected from GLCM matrix where pixels are treated as pair not individual. Basically occurrence matrix depends upon two important factors, relative distance *d* and orientation θ.

Here we show the different orientations (0°, 45 °, 90 °,and 135°) which will used in GLCM matrix.

$$P_{0^0,d}(I_1,I_2)=\left|\begin{Bmatrix}[(k,l),(m,n)]\in D\\ k-m=0,|l-n|=d\\ f(k,l)=I_1,f(m,n)=I_2\end{Bmatrix}\right|$$

$$P_{45^0,d}(I_1,I_2)=\left|\begin{Bmatrix}[(k,l),(m,n)]\in D\\ (k-m=d,l-n=-d)\vee(k-m=-d,l-n=d)\\ f(k,l)=I_1,f(m,n)=I_2\end{Bmatrix}\right|$$

$$P_{90^0,d}(I_1,I_2)=\left|\begin{Bmatrix}[(k,l),(m,n)]\in D\\ k-m=0,|l-n|=0\\ f(k,l)=I_1,f(m,n)=I_2\end{Bmatrix}\right|$$

$$P_{135^0,d}(I_1,I_2)=\left|\begin{Bmatrix}[(k,l),(m,n)]\in D\\ (k-m=d,l-n=d)\vee(k-m=-d,l-n=-d)\\ f(k,l)=I_1,f(m,n)=I_2\end{Bmatrix}\right|$$

From the above four sets, (k, l) = pixel point, f (k, l) = intensity at pixel point ($M\times N$) = image order and the ($M\times N$)×($M\times N$)= order of the image matrix. There are two figures, first represents the matrix of gray levels and second represents the gray level co-occurrence matrix of particular ROI of the image. Here we extract the 13 features in all four directions with offset value 1.

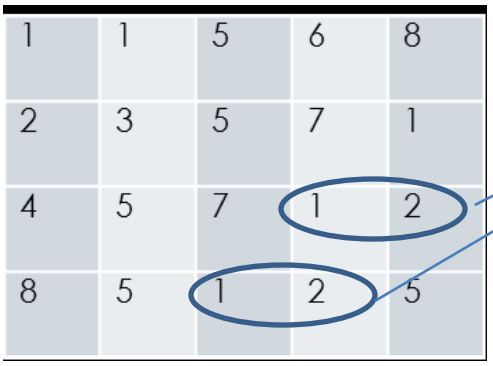


Fig-4.3 Matrix of Gray levels

	1	2	3	4	5	6	7	8
1	1	2	0	0	1	0	0	0
2	0	0	1	0	1	0	0	0
3	0	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0	0
5	1	0	0	0	0	1	2	0
6	0	0	0	0	0	0	0	1
7	2	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0

Fig-4.4 Co-occurrence matrix (GLCM)

By the Fig-4.3 is clearly represented that (1,2) pixel occur in two times in horizontal direction. So these numbers of recurrence are clearly described by GLCM matrix (Fig-4.4). The distance D which is also called offset and angle θ which is the direction orientation between two pixels are used for computation of the probability distribution. There are many types of offsets and orientations are defined as:

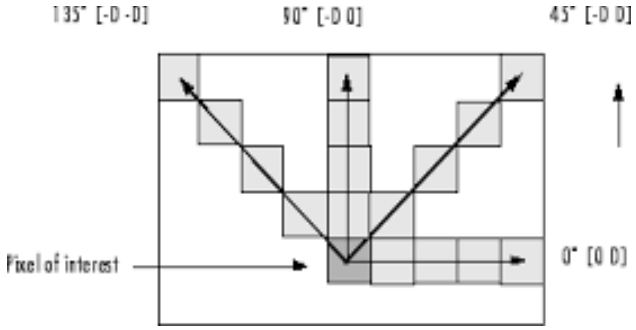


Fig-4.5 Different orientation

Table 4.2 GLCM features and their mathematical formulae [47]

S.No	GLCM Features	Mathematical Formulae
1	Angular second moment	$ASM = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\{P(i,j)\}^2$
2	Contrast	$CON = \sum_{n=0}^{G-1}(n)^2 \sum_{i=1}^{G}\sum_{j=1}^{G}\{P(i,j)\}, \lvert i-j\rvert = n$

3	Inverse difference moment	$IDM = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{1}{1+(i-j)^2} P(i,j)$
4	Entropy	$ENT = -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P(i,j) * \log\left(P(i,j)\right)$
5	Correlation	$COR = \sum_{i}\sum_{j} \frac{(i,j)P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
6	Sum of squares (variance)	$VAR = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \left(i-\mu\right)^2 P(i,j)$
7	Sum average	$SA = \sum_{i=2}^{2N_g} i.P_{x+y}(i)$
8	Sum variance	$SV = \sum_{i=2}^{2N_g} (i-f_4)^2 .P_{x+y}(i)$
9	Sum entropy	$SE = -\sum_{i=0}^{2N_g} P_{x+y}(i)\left\{\log\left(P_{x+y}(i)\right)\right\}$
10	Difference variance	$DV = \text{variance of } P_{x-y}$
11	Difference entropy	$DE = -\sum_{i=0}^{N_g-1} P_{x-y}(i)\left\{\log\left(P_{x-y}(i)\right)\right\}$
12	Information measures of correlation 1	$IMC1 = \frac{H_{xy} - H_{xy1}}{max\{H_x, H_y\}}$
13	Information measures of correlation 2	$IMC2 = \{1 - \exp[-2(H_{xy2} - H_{xy})]\}^2$

4.2.2.3 GRAY LEVEL RUN LENGTH MATRIX (GLRLM)

The basic of run-length matrix is proposed by researcher Galloway [54]. Coarse structure or long gray levels run are analyzed by gray level run length matrix. This matrix (GLRLM) is statistical based approach for higher runs of pixels. *RunLenght(m,n)* matrix has two parameters where *n*= total elements and *m*=intensity value [55].

$$M(\theta) = \left(\left(RunLength(m,n) \middle| \theta \right) \right), 0 \le m \le M_g, 0 \le n \le M_{\max}$$

Where, Ng is maximum gray level and $Mmax$ is the maximum length. If a pixel with high runs occurred in a specific direction will be considered as one run-length. Run-length matrix as shown in Figure 4.7

Fig 4.6 Image Matrix

Gray Level	Run Length (j)			
i	1	2	3	4
1	4	0	0	0
2	1	0	1	0
3	3	0	0	0
4	3	1	0	0

Fig 4.7 Run Length Matrix

Table 4.3 represents the extracted features from gray level run length matrix.

Table 4.3 GLRLM features and their mathematical formulae [54]

S.No.	GLRLM Features	Mathematical Formulae
1	Short run emphasis	$SRE = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{P(i,j)}{j^2}$
2	Long run emphasis	$LRE = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} P(i,j).j^2$
3	Gray-level non-uniformity	$GLN = \frac{1}{n_r} \sum_{i=1}^{M} \left(\sum_{j=1}^{N} P(i,j) \right)^2$
4	Run length non-uniformity	$RLN = \frac{1}{n_r} \sum_{j=1}^{N} \left(\sum_{i=1}^{M} P(i,j) \right)^2$
5	Run percentage	$RP = \frac{n_r}{n_p}$
6	Low gray-level run emphasis	$LGRE = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{P(i,j)}{i^2}$
7	High gray-level run emphasis	$HGRE = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} P(i,j).i^2$

8	Short run low gray-level emphasis	$SRLGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j)}{i^2.j^2}$
9	Short run high gray-level emphasis	$SRHGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j).i^2}{j^2}$
10	Long run low gray-level emphasis	$LRLGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j).j^2}{i^2}$
11	Long run high gray-level emphasis	$LRHGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}P(i,j).j^2.i^2$

4.2.2.4 STATISTICAL FEATURES MATRIX (SFM):

The basic of statistical feature matrix is proposed by C.M. Wu [56]. It is used for visual perception based texture feature extraction. Correlation between two statistical feature matrixes is simply defined by measurement of distance. In this study we applied this matrix to the categorization of 480 sampled DDSM mammogram images. This approach has shown the better result than statistical texture feature models. Here we have been extracted four features based on visual perception. Four statistical matrix features are:

a) COARSENESS:

Coarseness can be defined in terms of repetition rates and scale. A tissue image comprises textures at numerous scales. The main objective to study of coarseness is to classify the largest size in a micro structure tissue.

b) CONTRAST:

In visual observation, contrast is evaluated by the dissimilarity in the color and brightness of the image.

c) PERIODICITY:

Periodicity is the state or fact which have being frequently recurring or having periods in sample image.

d) ROUGHNESS:

Roughness is an element of surface texture in image. Roughness is measured by the variations in the direction of the typical vector of an actual surface from its perfect form.

4.2.2.5 LAW'S TEXTURE ENERGY MEASURES (LTEM)

Law's Textural Energy Measures (LTEM) is additional broadly famous method for texture based classification. LTEM method was developed by K. I. Laws [57] in 1980. There are many of local masks for textural detection in many ways as wave, level, spot, edge, and ripple. All masks have different length. When we convolved these local masks, different types of masks generated in result. These new masks are so useful for extracting the histogram based features. There are three types of local masks of length 3, which are defined as:

$$L3 = [1 \quad 2 \quad 1];$$
$$E3 = [-1 \quad 0 \quad 1];$$
$$S3 = [-1 \quad 2 \quad -1];$$

A total of (6×5=30) features extracted from convoluted masks to each other out of these mask vectors.

Masks for the length 5 are represented as:

$$L5 = [1 \quad 4 \quad 6 \quad 4 \quad 1] \quad \text{(Level)}$$
$$E5 = [-1 \quad -2 \quad 0 \quad 2 \quad 1] \quad \text{(Edge)}$$
$$S5 = [-1 \quad 0 \quad 2 \quad 0 \quad -1] \quad \text{(Spot)}$$
$$W5 = [-1 \quad 2 \quad 0 \quad -2 \quad 1] \quad \text{(Window)}$$
$$R5 = [1 \quad -4 \quad 6 \quad -4 \quad 1] \quad \text{(Ripple)}$$

A total of (15×5=75) features extracted from length 5 local masks by convoluted masks.

Masks for the length 7 are represented as:

$$L7 = [1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1]$$
$$E7 = [-1 \quad -4 \quad -5 \quad 0 \quad 5 \quad 4 \quad 1]$$
$$S7 = [-1 \quad -2 \quad 1 \quad 4 \quad 1 \quad -2 \quad -1]$$

A total of (6×5=30) features extracted from length 5 local masks by convoluted masks.

4.2.2.6 GABOR WAVELET FEATURES:

Dennis Gabor was developed Gabor function in 1946 [58]. Wavelets transform and Gabor Filter based on multi-channel or multi-resolution examination. So these models are more important for texture related applications like segmentation or classification. Gabor filters have multiscaled functions which are used for the feature extraction and also image can be examined with magnification. Gabor wavelets have energy capturing capability at a particular scale and orientation. Gabor functions have two types based on the dimensionality, which are:

a) ONE-DIMENSIONAL GABOR FILTER FUNCTION

Gabor function is the product of composite sinusoidal and a Gaussian functions. It is also called as harmonic oscillator. The mathematical formula for one dimensional Gabor filter function is defined below, where W = modulation frequency, ρ = standard deviation and σ = scale of Gaussian function.

$$G(x) = \frac{1}{\sqrt{2\pi\rho}} exp\left(\frac{-x^2}{2\sigma^2}\right) exp\left(j\pi Wx\right)$$

b) TWO-DIMENSIONAL GABOR FILTER FUNCTION

A common 2-D Gabor filter is basically a sinusoidal function. Two dimensional Gabor function is defined as

$$GB(x, y) = g_\sigma(x, y) exp\left[2\pi jW\left(x\cos\theta + y\sin\theta\right)\right]$$

Where

$$g_\sigma(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right]$$

The parameter θ signifies the direction parameter, σ signifies the scale, and W is the modulation frequency of the Gaussian function. These types of Gabor functions have many of application like tuning, image compression and classification for texture. When an image $(x,)$ is convoluted with a Gabor function $(x,)$, the result gives a Gabor filtered image.

The Gabor wavelets are limited with in two-dimensional Gaussian envelope. There are many direction or orientation and scale which are similar to the complex planar. Every Gabor wavelet comprises a specific orientation and wavelength. The Gabor wavelets are usually measured as

distinct set of self-similar functions. If (x,) is the mother Gabor wavelet, at that point this self-similar filter group is acquired by appropriate rotations and dilations of (x,). The Discrete Gabor Wavelet Transform (DGWT) of image (x,) size ($M \times N$) is termed as:

$$G_{pq}(x,y) = \sum_{s}\sum_{t} A(x-s, y-t)\psi^{*}_{pq}(s,t)$$

ψ^{*}_{pq} is complex conjugate of ψ_{pq}, which is formed by the spin and dilation of mother wavelet ψ

$$\psi(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right] exp(2\pi jWx)$$

Here show the generating functions which constructs the Gabor Wavelet function, defined as:

$$\psi(x,y) = a^{-p}\psi(\bar{x},\bar{y})$$

$$\bar{x} = a^{-p}(x\cos\theta + y\sin\theta)$$

$$\bar{y} = a^{-p}(y\cos\theta - x\sin\theta), for\ a > 1 \text{and}\ \theta = \frac{q\pi}{Q}$$

Where a=scale factor, P= total no of scales, Q=total no of orientations

In the present study 24 GWT is developed by settings of angle and scale. We have 40 ROIs of one class. So 40 Gabor filtered output images are created by convolution between an image and these GWT filters. These Gabor filtered outputs are used in the texture classification based application. These output Gabor filtered images comprise diverse textural information for classification purpose. The essential energy parameter which is generated by Gabor filtered outputs, is described below

$$EN_{GWT}(p,q) = \sum_{x}\sum_{y} G_{pq}(x,y)$$

Standard deviation σ_{GWT} and mean μ_{GWT} and for every Gabor filters are defined below. These standards indicate the feature of an identical texture image.

$$\mu_{GWT} = \frac{EN_{GWT}(p,q)}{MN}$$

$$\sigma_{GWT} = \sqrt{\frac{\sum_{x}\sum_{y}\left|G_{pq}(x,y)\right| - \mu_{pq}}{MN}}$$

4.2.3 FEATURE SELECTION

The technique for the optimal features selection is entering out as ample of unrelated, frequent and redundant information as probable. Basically it is the part of pattern recognition. When we used feature selection learning procedures, it is found that this method is more efficiently else other would be composite and hectic [46]. This method chooses the best features among all complex features by decreases the dimensions of the dataset. It can be improved the performance of the classification task and easily interpreted the representations.

The researcher Genari *et al.* concluded that texture attributes are related if their standards differ methodically with class relationship. In other words, a feature is correlated or predictable in the class it is useful; otherwise it is unrelated [59].

If many of other features are highly predictive with feature then that feature is supposed to be redundant. r_{zc} is the relationship among the summed components and the external variables.

$$r_{zc} = \frac{kr_{zi}}{\sqrt{k + k(k-1)r_{ii}}} \qquad \text{(Equation 4.1)}$$

Where r_{zi} = mean of the correlations among the components and the external variable, k = amount of components, , and r_{ii} = mean inter-correlation between modules.

4.2.3.1 CORRELATION FEATURE SELECTION

Correlation Feature Selection (CFS) is simple working filter procedure that ranks feature subsections conferring to a correlation based heuristic assessment function. The result of the accuracy assessment function consists of subgroup of features. Those optimal features are highly predictive with that class. The accuracy assessment task is same as equation 4.1 with marginally changed, that is

$$CFS_s = \frac{mi_{cf}}{\sqrt{m + m(m-1)i_{ff}}} \qquad \text{(Equation 4.2)}$$

Where CFSs = heuristic "merit" of a feature subsection S

$i_{cf=}$ external correlation between features and class

i_{ff} = internal correlation between feature and feature

The flowchart of CFS based technique shown below:

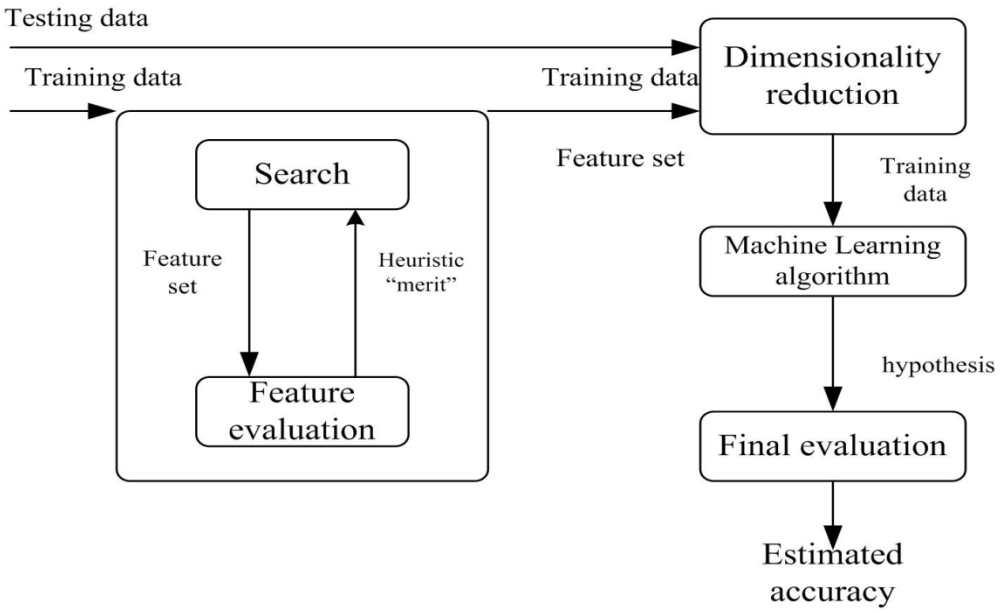
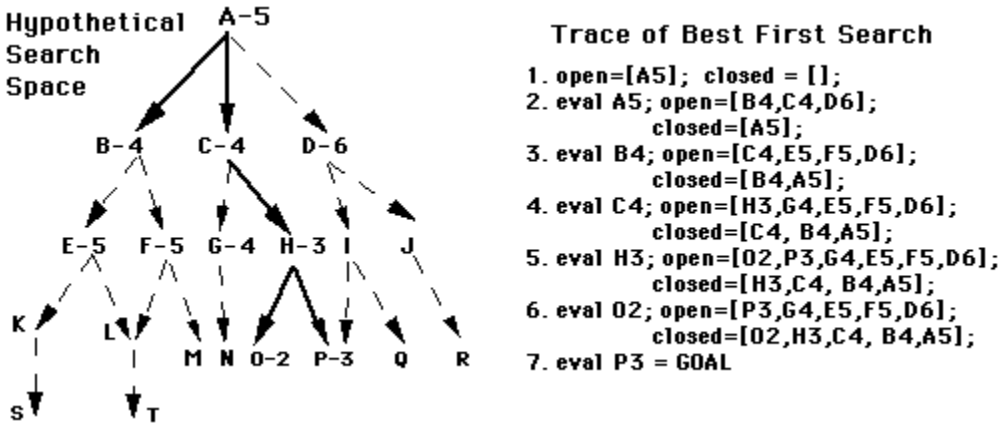


Fig 4.8 Correlation based feature selection flowchart

In present study best first search algorithm is used for correlation based feature selection method. Best first search is a examine procedure which discovers a graph by increasing the most likely node selected conferring to specified rule. The algorithm for best search is:



4.2.4 CLASSIFICATION MODULE

In present study three classifiers are used for classification purpose: Support Vector Machines (SVM), Multilayer Perceptron (MLP) and k-Nearest Neighbors (k-NN).

1. SUPPORT VECTOR MACHINE

SVM can be performed for dual category problems. It categorizes the complex data by defining the finest hyper plane. The hyper plane distributes the plane into two classes the one side of the plane belongs to one class and second side of the plane belongs to other class. Larger width margin boundary gives the best hyperplane between classes

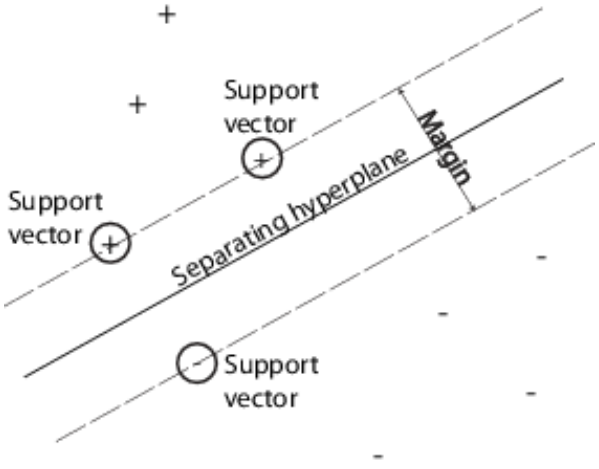


Fig 4.9 Basic of Support vector machine [60]

Here Figure 4.9 represents the basic methodology of support vector machine. There are two classes. First class is represented by + sign, and second class represented by – sign. The vectors which are closest to the hyper plane and lies in the boundary called support vectors. [60]
There are many parameters used for categorization.

- ∈ (For round-off error)= 1.0E-12
- Kernel= Linear kernel
- C (Soft margin constant)= 1

2. MULTILAYER PERCEPTRON:

MLP is encouraged by the human brain. There are three basic layers in multilayer perceptron model. First layer is called input layer, last layer is called output layer and middle layer which connect the input and output layer is called hidden layer. MLP is one case of feed-forward neutral network. It associates numerous perceptron (simple neural networks) to generate non-linear decision boundary. Figure 4.10 shows the layers including in multilayer perceptron model. Numbers of hidden layers depend upon complexity of the system. In MPL model, every

perceptron accepts a response (i.e. stimuli) for processing and delivers the output via its connected relations to the adjacent perceptron.

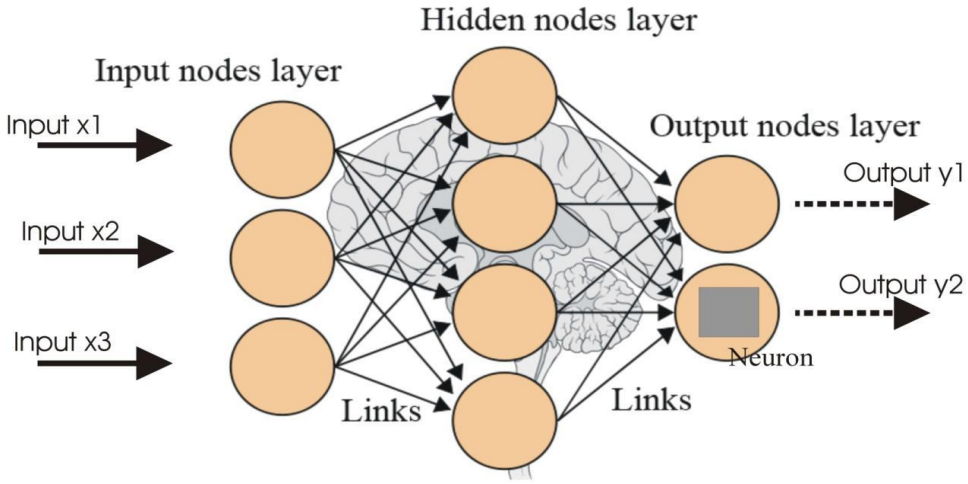


Fig 4.10 Multilayer perceptron model [61]

There are four main sections in ANN,

1. Hidden interconnections between input and outputs,

1. For getting a stimuli, elements of the perceptrons are stimulates.

3. Learning function is used for calculating the weights between input and output.

4. An activation function which used to transform input into output confidential the perceptron.

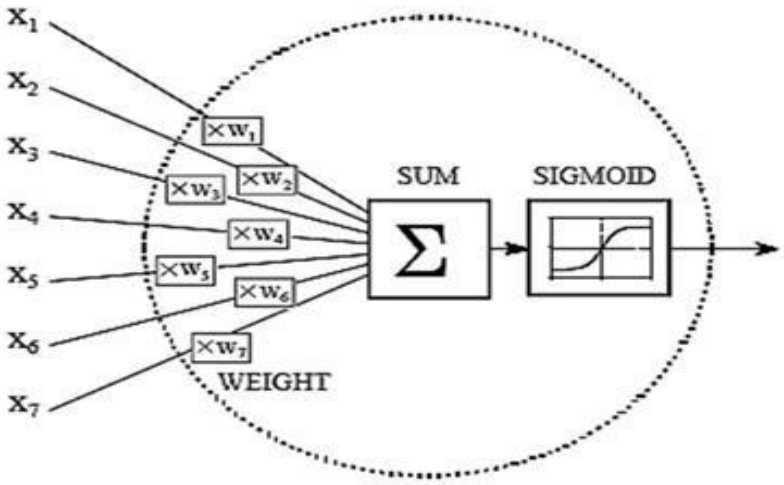


Fig 4.11 Multilayer perceptron internal structure

Figure 4.11 defines the interior construction of the perceptron which has input Xi, input with weight XWi and Sigmoid as Activation function.

There are many of parameters used for classification:

- Momentum =0.2
- Learning rate =0.1
- Validation threshold=20
- Activation function: Sigmoid
- Training time =500

3. K-NEAREST NEIGHBOURS

In K-NN, classification task is completed based on the type of simply one adjacent neighbor so the method is recognized as nearest neighbor classifier. If there are more than one neighbor this method is known as k-Nearest Neighbors.

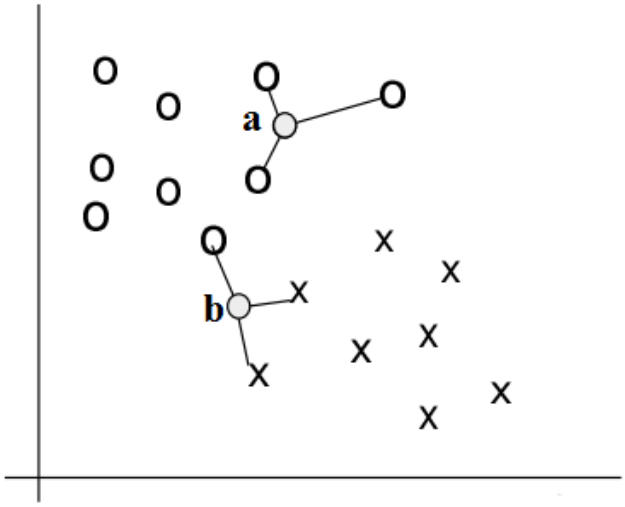


Fig 4.12 Basic example of k-NN

Here figure 4.12 defines the 3- nearest neighbor classifier problem, which will be solved by k-NN classifier. In above problem due to nearest neighbor ‘a’ is existing as same class ‘O’ but ‘b’ has some problem, because it has two adjacent neighbor ‘X’ and ‘O’. This type of problem have examined by majority of feature points. It is solved by K-NN conveniently. In last it is concluded that ‘b’ involves to class ‘X’.

There are some parameters which include in k-NN:

- K=1
- Search procedure: Linear nearest adjacent search process.
- Distance function: Euclidean distance

4.3 PROPOSED METHODOLOGY (ALGORITHM 2)

There are many of reason behind for increase malignant tissues. Masscalcification in the breast is one of them. Calcifications are tiny clusters of calcium. That spots are the high intensity variations in the breast. Micro-calcification may be unrestrained and found stationary in a mass. Typically the additional cell development is not tumorous, but sometimes constricted clusters of micro-calcification can show initial breast cancer. Dispersed micro calcifications are typically a symbol of benign breast malignancy. 80% of the micro-calcification is nonthreatening. Micro-calcification in the breast indicate as white spots on breast x-rays. A more broad review on recognition and classification approaches of micro calcifications can be originate in [62]

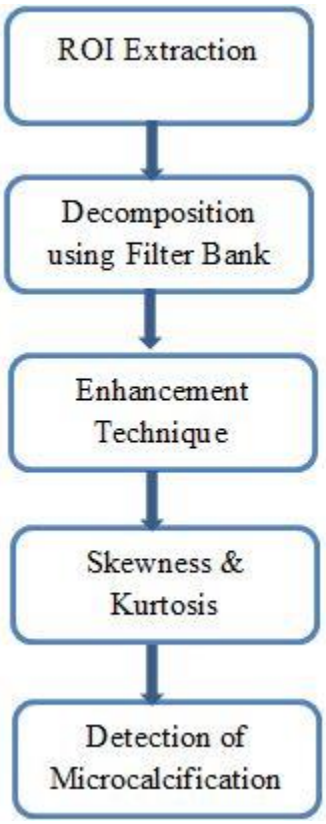


Fig 4.13 Algorithm for micro calcification detection

4.3.1 ROI EXTRACTION

In algorithm 2, firstly we extract the 256×256 size ROI manually from the center of the breast as we cover most dense and tiny calcium clusters area of the breast. The whole process of ROI extraction is done on MATLAB software.

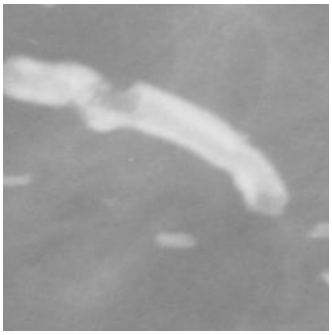


Fig 4.14 256×256 size ROI

4.3.2 DECOMPOSITION USING FILTER BANK

In mammograms for the classification in between normal tissue and micro calcification clusters, it is necessary to identify both linear components, like mammary ducts, blood vessels etc. and nodular components, like micro calcifications. Generally second derivative is employed for the detection of these components. For the nodular structure, the second derivatives values tend to become negative in all directions. Conversely, for the linear structure the second derivative value tends to zero in the direction of linear structure axis, while it tends to negative in the direction which is perpendicular to the linear structure axis. Hence, the second derivative filters can be employed for the identification as well as enhancement of the linear structure and the nodular structure. A Min-DD Filter (minimum directional difference) was developed by Shimizu *et al.* [62]. It depends upon the minimum second derivatives value in whole direction. A Max-DD Filter (maximum directional difference), which is depends upon the maximum second derivatives value in whole directions. For detection of large lung nodules in chest x-ray imaging technique, they used Min-DD Filter.

As we discussed in the preceding section, the linear component and the nodular component can be distinguished by means of second derivative value. The filter length remains constant for the

second derivative method [63] of Shimizu even though lung nodules have numerous sizes. Hence, detection of linear as well as nodular structures of several sizes becomes possible with more accuracy using second derivative filters. Since second derivative is generally influenced by noise, hence it is essential to shape properly linear as well as nodular structure with a smoothing operator. However, by using filter banks, these kinds of issues are easily resolvable, which contains of low-pass and high-pass filters of different lengths. After detection of micro calcification cluster, the next issue is to identify whether the lesion is malignant or benign. For differentiating in between malignant and benign micro calcifications most of the investigators have developing computerized based automated schemes [57]. However, image features like size, irregularity are generally used to find out the possibility of malignancy. Hence, in these schemes, it is necessary to enhance micro calcifications while keeping their shapes.

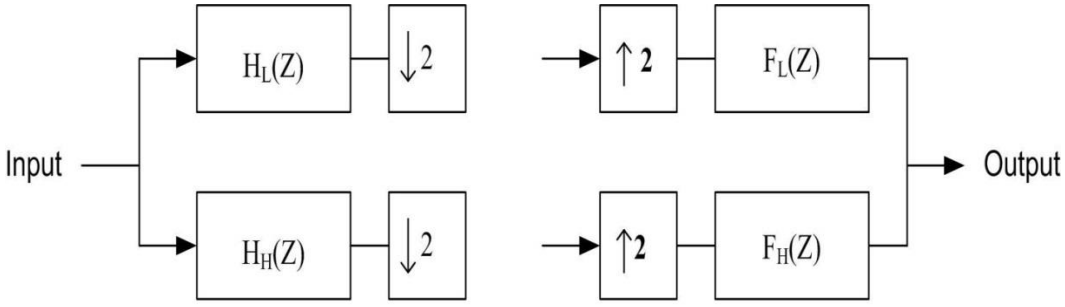


Fig 4.15 2-channel decomposition filter

Fig. 4.15 displays a two-channel decomposition filter bank. The left part of decomposition filter also called analysis bank consists of high pass filter, low pass filter and down sampling operator (↓) which eliminates the ripples after filtering. The right part synthesis bank has a low-pass filter, a high-pass filter, and an up sampling operator (↑) which introduces a zero in the odd components. The filters of two channel filter bank must fulfill this condition

$$H_L(-Z)\, F_L(Z) + H_H(-Z)\, F_H(Z) = 0$$

4.3.3 ENHANCEMENT TECHNIQUE

Wavelet transform is used for improvement in contrast enhancement. Wavelet transform included the 2-level decomposition filter bank. avelet and filter bank combined components gives the low pass components and high pass components. From wavelet-transform, these procedures can be employed without using two factors down sampling. In this method, high pass

variables will be blocked if the value is smaller than the threshold value and will be amplified if the value is larger than threshold. To achieve this non-linear type of operation, the formula used can be given as:-

$$f(y) = a[sigm(c(y-b)) - sigm(-c(y+b))]$$

Where

$$a = \frac{1}{sigm(c(1-b)) - sigm(-c(1+b))} \quad 0<b<1$$

$$sigm(y) = \frac{1}{1+e^{-y}}$$

4.3.4 DETECTION OF MICROCALCIFICATION

The microcalcification recognition procedure was recommended by using first order statistical approaches for example skewness and kurtosis. Skewness measures the asymmetry and kurtosis measures impulsiveness of the distribution, defined as

$$\text{Skewness } \gamma_1 = \frac{E[(x-E(x))^3]}{E[(x-E(x))^2]^{3/2}}$$

$$\text{Kurtosis } \gamma_2 = \frac{E[(x-E(x))^4]}{E[(x-E(x))^2]^2} - 3$$

If an area holds micro-calcification then, the regularity of the distribution of image constants is demolished due to the impulsive nature of micro-calcification. Kurtosis also measures peakness of the distribution. And also check tails of the distribution heavier or not. Thus skewness and kurtosis features are efficient way to detect micro calcification spots with asymmetric and high peak tailed distributions**.**

$$\tau_x = \begin{cases} 0 & \gamma_1 < T_1 \text{ or } \gamma_2 < T_2 \\ 1 & \gamma_2 \geq T_1 \text{ or } \gamma_2 \geq T_2 \end{cases}$$

This type of problem based on assumption type testing problem, where T_1 and $\boldsymbol{T_2}$ are thresholds value of skewness and kurtosis, correspondingly. 0 represents there is no microcalcification in the area, and 1 represents there is microcalcification in the area.

CHAPTER 5

RESULTS AND DISCUSSION

This chapter shows with the results of the work at various levels. The results are presented as:

- Texture Feature Extraction
- Dimensionality reduction
- Evaluate the accuracy of performance.
- Detection of micro calcification using wavelet.

5.1 FEATURE EXTRACTION

Texture features extracted from three lesions of mammograms; i.e. normal, benign, and cancer. Different texture models are used for classification purpose. From the ROIs total 242 texture features are extracted from different texture models. Table 5.1 shows the overview of texture features:

Table 5.1 An overview of total features

S.No.	FEATURE NAME	FEATURE TYPE AND ORIENTATION	NUMBER OF FEATURES
1.	FOS	Intensity based features	6
2.	GLCM	Texture based feature at 0°,45°,90°,135° with offset value '1'	13×4=52
3.	GLRLM	Texture based feature at 0°,45°,90°,135°	11×4=44
4.	SFM	Visual perception based features	4
5.	LTEM	Texture based features with 3, 5, 7 length filters. From convolved images mean, std deviation, skewness, kurtosis and entropy are calculated.	[6+15+6] ×5=135
6.	GABOR WAVELETS	Texture based feature at 6 different angles (30°, 60°, 90°, 120°, 150°, 180°) with 4 scales. From these convolved images mean, std deviation, skewness, kurtosis and entropy are calculated.	5
	TOTAL FEATURES		242

5.2 DIMENSIONALITY REDUCTION

It has been stated before, a huge amount of features make the data study more complex, confused and takings additional time; therefore feature selection process have been used for dimensionality reduction. In this work, different optimal features are extracted from correlation base feature selection (CFS) technique with best search algorithm. Total 53 optimal features are selected from CFS method. Here table 5.2 has shown the optimal features.

Table 5.2 Optimal features selected by CFS

Texture models	**Optimal features**	
FOS	Mean, skew-ness, kurtosis	
GLCM	0º	Sum_avg, diff_var, diff_ent, Inf corr1
	45º	Corr, Sum_avg, diff_var,diff_ent, Inf corr1
	90º	Sum_avg, diff_var, Inf corr1
	135º	Cont, Corr,Sum_avg , diff var, diff ent
GLRLM	0º	SRE, GLN, LGRE, LRLGE
	45º	LRE, LGRE, LRLGE
	90º	SRE, HGRE, LGRE
	135º	LRE, RLN, SRHGE
LTEM	d=3	LL_mean, EE_std, EW_std, LE_skew
	d=5	LL_mean, LL_std, SS_std,LW_std,EW_std, ER_std, RS_Std, WW_skew, WS_kurt, EW_ent
	d=7	LL_std, EE_mean, LL_skew
SFM	Contrast	
Gabor Filter	Mean, Kurtosis	

5.3 EVALUATION OF PERFORMANCE

The classifiers were used on all the texture features, to check and relate their correctness of classification for the similar database. The results are showed in the following experiments.

5.3.1 EXPERIMENT 1

In experiment 1, the attempt has been made to select the best ROI size from different ROIs, which have best performance among all. There are total 480 ROI images used for this purpose. For estimating the performance, 10-Fold cross validation has been used. In all iterations i.e., two-third of the data is selected as training and remaining one third is used for testing. The overall accuracy is calculated by the average accuracy of different texture models. Table 5.3 represents the accuracy corresponds to all classifier for all size of ROIs.

Table 5.3 Classification performance of texture models for different ROIs

ROI Size	**Classifiers**	**Accuracy by different texture models (%)**												
		FOS	**GLCM**				**GLRLM**				**SFM**	**LTEM**	**Gabor**	**Average**
			0º	**45 º**	**90 º**	**135º**	**0º**	**45 º**	**90 º**	**135 º**				
	KNN	68	65	66	65.5	67	63	63.5	66.3	68.9	67	60	53	**64.43**
64×64	SVM	62	49	55	43.4	60	47	58	54.6	62	63	46.8	49	**54.15**
	MLP	65	63	64.2	65	69.8	63	60	64.2	71.1	69	62	43	**63.27**
	KNN	67.5	66.5	67	70	69.5	64	67	70.7	62.4	70	65	65	**67.05**
128×128	SVM	64	66	63.4	61.7	63	60	63.4	60.7	66.8	72	70.9	61	**64.40**
	MLP	68	72	70	74	76	73	70	73.2	75	72.6	72	63.5	**71.60**
	KNN	75	74	72.8	75	78	75	73.5	79	84	85	78	75	**77.02**
256×256	SVM	73.5	69	69	71.7	75	72	76	73.6	78	82	76.7	73.5	**74.16**
	MLP	79	81	82.8	82	84.5	79.5	81	84	85	87.8	88	78	**82.71**
	KNN	73	60	65	63	65.8	64	61	67	67.5	67	67	69	**65.77**
512×512	SVM	70	58.4	60	63.4	65	55	60	62.6	63.6	73	53.4	55	**61.61**
	MLP	74	68	65	67	72	70	68	65	72.8	75	70	64	**69.23**

The experiment 1, we have extracted the many texture features from numerous texture models (FOS, GLCM, GLRLM, SFM, LTEM and Gabor Filter). Here we study the performance of classification. The parameter for performance is taken as accuracy. Accuracy is calculated by using different type of machine learning classifiers. Figure 5.1 represents the accuracy corresponding to all texture models for different ROIs. With the help of classification performance we find out the appropriate size of ROI which has best accuracy among all.

CONCLUSION OF EXPERIMENT-1

It has been concluded that squared shaped 256×256 ROI size has best classification performance among all ROI size and best location for ROI extraction is center part of the image. It shows 82.71% accuracy by MLP classifier, 77.02% accuracy by K-NN classifier and 74.16% accuracy by SVM classifier. By experiment 1, we prove that 256×256 size ROI has best accuracy from all of the classifiers. So we choose 256×256 size ROI for next experiments.

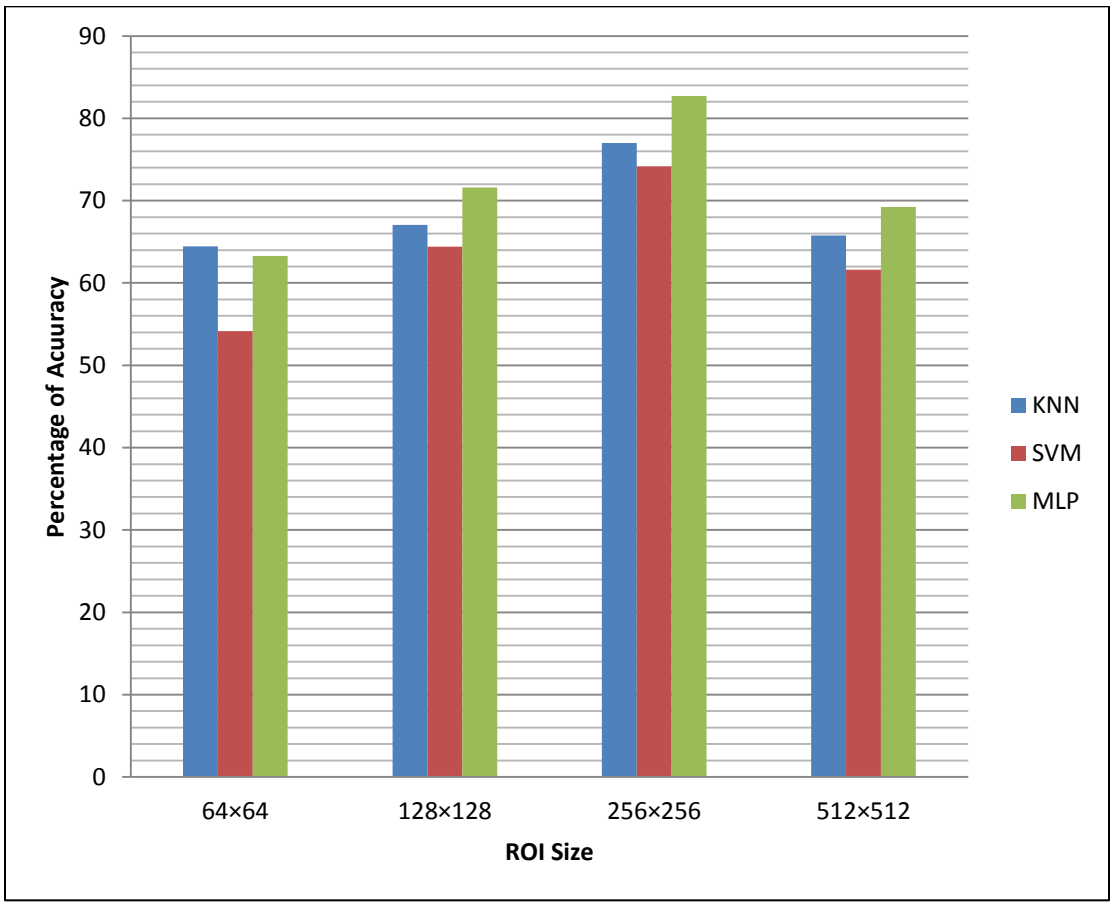


Fig 5.1 Classification performance of different ROIs

5.3.2 EXPERIEMENT 2

In experiment 2, we discussed about performance of classification when optimal features are used as input vector. Here we choose 256×256 ROI size for selected the optimal features from correlation based feature selection method using best first algorithm. Table 5.4 has shown the classification performance of individual and combined optimal features.

Table 5.4 Classification performance of individual and combined optimal features (256×256)

Texture models		**Classifiers**	**Accuracy (%)**
FOS		k-NN	68.33
		SVM	71.67
		MLP	**76.67**
GLCM	0º	k-NN	70.83
		SVM	67.5
		MLP	**79.16**
	45º	k-NN	72.5
		SVM	67.33
		MLP	72.5
	90º	k-NN	74.5
		SVM	69.16
		MLP	76.67
	135º	k-NN	70.83
		SVM	67.89
		MLP	73.33
GLRLM	0º	k-NN	72.5
		SVM	73.34
		MLP	**80.83**
	45º	k-NN	65.83
		SVM	69.16
		MLP	75.83
	90º	k-NN	68.34
		SVM	74.16
		MLP	75
	135º	k-NN	72.80
		SVM	73.67
		MLP	76.76
SFM		k-NN	78
		SVM	74.5
		MLP	**85**

LTEM	d=3	k-NN	85
		SVM	81.67
		MLP	**86.67**
	d=5	k-NN	83.34
		SVM	83.34
		MLP	85.93
	d=7	k-NN	74.5
		SVM	69
		MLP	76.67
GABOR Filter		k-NN	69
		SVM	73
		MLP	**75**
Selected features in combinations		k-NN	79.16
		SVM	84.16
		MLP	**90**

The experiment 2 has investigated the accuracy after CFS method. Here we have selected 53 optimal texture features from various texture models like FOS, GLCM, GLRLM, SFM, LTEM and Gabor Filter. Here accuracy is calculated by optimal features as individually and combined.

CONCLUSION OF EXPERIMENT-2

It has been concluded that combination of all selected optimal features have best classification performance. It shows 90% accuracy by MLP classifier, 84% accuracy by SVM classifier and 79% accuracy by K-NN classifier. If we have taken individual optimal features then accuracy is not vary show much, it is almost same as previous accuracy when CFS method not used. By experiment 2, we have proven that combined optimal features have best accuracy from all of the classifiers. So features selection method is very useful for enhance the performance.

The result shows that when combined optimal features are used then accuracy enhanced. MLP classifier has best accuracy among all classifier for this work. Table 5.5 to table 5.10 has shown the confusion matrix, when individual and combined optimal features are used as input vector to the multilayer perceptron classifier.

Table 5.5 Confusion matrix for FOS optimal features

Actual Classes	Predicted Classes		
	NORMAL	BENIGN	CANCER
NORMAL	34	4	2
BENIGN	3	31	6
CANCER	3	10	27

Table 5.6 Confusion matrix for GLCM ($\theta = 0°$) optimal features

Actual Classes	Predicted Classes		
	NORMAL	BENIGN	CANCER
NORMAL	36	1	3
BENIGN	2	31	7
CANCER	2	10	28

Table 5.7 Confusion matrix for GLRLM (($\theta = 0°$) optimal features

Actual Classes	Predicted Classes		
	NORMAL	BENIGN	CANCER
NORMAL	33	2	5
BENIGN	0	34	6
CANCER	1	9	30

Table 5.8 Confusion matrix for SFM optimal features

Actual Classes	Predicted Classes		
	NORMAL	BENIGN	CANCER
NORMAL	39	1	0
BENIGN	0	33	7
CANCER	0	10	30

Table 5.8 Confusion matrix for LTEM (d=3) optimal features

Actual Classes	Predicted Classes		
	NORMAL	BENIGN	CANCER
NORMAL	40	0	0
BENIGN	0	34	6
CANCER	0	10	30

Table 5.9 Confusion matrix for Gabor Filter optimal features

Actual Classes	Predicted Classes		
	NORMAL	BENIGN	CANCER
NORMAL	33	5	2
BENIGN	3	29	8
CANCER	2	10	28

Table 5.10 Confusion matrix for combined optimal features

Actual Classes	Predicted Classes		
	NORMAL	BENIGN	CANCER
NORMAL	34	2	4
BENIGN	1	37	2
CANCER	0	3	37

5.4 DETECTION OF MICRO CALCIFICATION

5.4.1 EXPERIMENT-3

There are some steps which carried out during implementation and simulation.

1. There are 120 mammograms of DDSM database, which are used as test mammograms for this study. The size of mammogram is 256×256.
2. We used Haar wavelet basis function with four-level decomposition filter bank. Here we don't use the 2-lavel down sampling. We used the wavelet transform for sustain the dimension of mammograms. The other advantage of wavelet transform is to reduce lost information.
3. In four levels decomposition filter bank, if high pass value is less than threshold hence high pass components will be suppressed otherwise enhanced the high pass components.
4. Microcalcification detection is done by the basis of skewness and kurtosis.
5. The simulation result of micro calcification detection has been shown below.

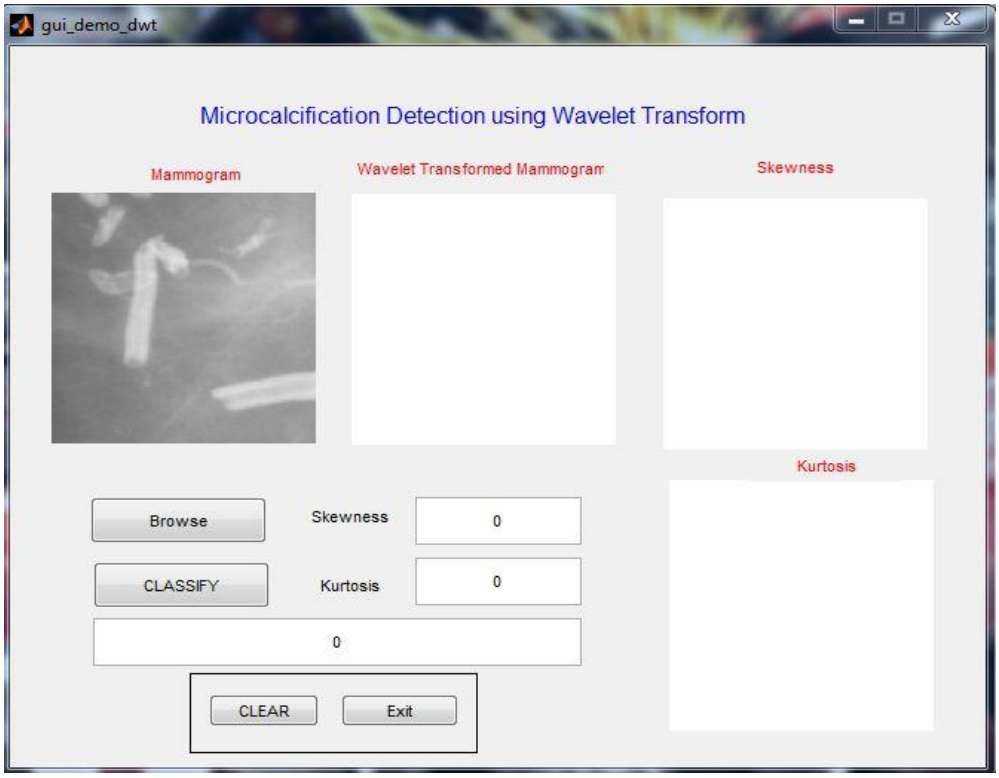


Fig 5.2 GUI for micro calcification detection (Select the ROI)

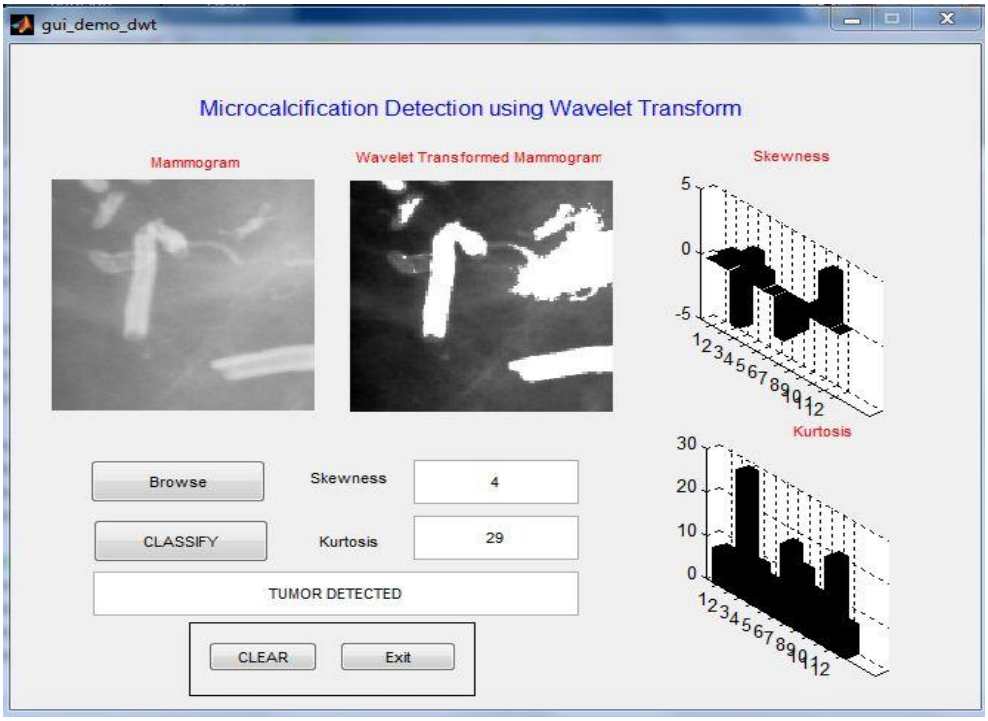


Fig 5.3 GUI for micro calcification detection (classify using wavelet transform)

CONCLSION OF EXPERIMENT-3

Here we made an efficient GUI model for detect the breast tumor in early stage. The regions of clustered microcalcification can be covered and the occurrence additional place of clustered microcalcification could be measured to simplify the analysis. From the 120 test images, there were the 112 test images result a good detection process and just 8 images was fail. It shown 93.34% success rate. The advantage of this algorithm is that processing is simple and does not require a full decomposition and reconstruction.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6.1 CONCLUSION

This dissertation represents the classification of mammograms using intensity and texture analysis. 480 mammograms are classified into normal, benign and cancer breast tissue type based on breast mass or density. Total 242 features are extracted for texture analysis for this work. 53 optimal features are selected from CFS method out of 242 features. K-Nearest Neighbor, Support Vector Machine and Multilayer Perceptron used for classification task.

Conclusions of the work are listed below:

(i). All the four ROIs, 256×256 size ROI has best accuracy by all of the classifier.

(ii). When categorizing the mammograms into three classes (normal, benign and cancer), the best results having accuracy of 82.71 % with MLP classifier (LR=0.1).

(iii). When classifying the mammograms into three classes, combined optimal features give the best results having accuracy of 90 % with the classifier MLP (LR=0.1).

(iv). We effectively made a GUI model for detect the microcalcification using filter bank and wavelet transform with 93.34% accuracy.

It shows that optimal features effectively used for classification of mammograms. The classifiers used were best data mining algorithms; hence they are more reliable too. Confusion matrix also shows that cancer breast tissue can be effectively classified from malignant tissue which is an indicator of having the chances of cancer.

6.2 FUTURE SCOPE

Even though it is proved that texture analysis is a very effective tool for classifying the mammograms into various classes, it does not differentiate completely. Some of the future scopes are listed below:

(i). Using different approaches like naïve Bayes, random forest, advanced computational techniques etc. the accuracy of classifying the mammograms into three classes can be improved.

(ii). Benign and cancer type breast tissue look alike sometimes as the malignant type breast tissue is developing stage towards the dense type. So sometimes it is difficult to differentiate between them. Some new methods can be employed/developed to differentiate the two effectively.

(iii). Try to make GUI model for whole CAD system.

LIST OF PUBLICATIONS

- Paper title "Investigation on ROI Selection for Mammograms Using Texture Models and Machine Learning Classifiers" presented in Shannon 100- 3rd International Conference on Computing Science organized by Lovely Professional University 8-9th April-2016, which is in proceeding of Indian Journal of Science & Technology (H-Index Scopus Journal- SJR-1.3)

- Dharmesh Singh and Mandeep Singh, "Classification of mammograms using Support Vector Machine", International Journal of Signal Processing, Image Processing and Pattern Recognition (IJSIP),Vol. 9, No.5(2016),pp.259-268.
  URL: http://dx.doi.org/10.14257/ijsip.2016.9.5.23

- Dharmesh Singh and Mandeep Singh, "Breast tumors detection: Gabor wavelet vs. statistical features", communicated to Medical & Biological Engineering & Computing (Springer).

REFERENCES

[1] Population Based Cancer Report http://ncrpindia.org/Annual_Reports.aspx. Date: 01/02/2016

[2] Cancer facts and figure. http://www.cancer.org/research/cancerfactsstatistics/Cancerfacts figures2015, Date: 01/02/2016

[3] Cheng H.D. and Cui M., "Mass lesion detection with a fuzzy neural network," *Pattern Recognition*, vol. 37, no. 6, pp. 1189-1200, 2004.

[4] Tabar L., Fagerberg G., Duffy S., Day N., Gad A. and Grontoft O., "Update of the Swedish two-country program of mammographic screening for breast cancer," *The Radiologic Clinics of North America*, vol. 30, pp. 187-210, 1992.

[5] Kopans D., *Breast Imaging* (cd 2), Philadelphia: J.B. Lippincott Company, 1989.

[6] Howard J., "Using mammography for cancer control: an unrealized potential," *CA: A Cancer Journal for Clinicians*, vol. 37, pp. 33-48, 1987.

[7] Cheng H.D., Lui Y.M., and Freimanis R.I., "A novel approach to microcalcification detection using fuzzy logic technique," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 442-450, June 1998.

[8] Mousa R., Munib Q., and Moussa A., "Breast Cancer Diagnosis System based on Wavelet Analysis and Fuzzy-Neural," *Expert Systems with Applications*, vol. 28, pp. 713-723, 2005.

[9] Sampat PM., Markey MK, and Bovik AC., "Computer-aided detection and diagnosis in mammography," *Handbook of Image and Video Processing, 2nd ed., A. C. Bovik Ed. Academic Press*, pp.1195-1217, 2005.

[10] Manjunath B., Ma W., "Texture feature for browsing and retrieval of image data", *IEEE Trans PAMI*, pp. 837-842, 1996.

[11] Christoyianni I.,Koutras A., Dermatas E., and Kokkinakis G., "Computer aided diagnosis of breast cancer in digitized mammograms," *Computerized Medical Imaging and Graphics*, vol. 26, pp. 309–319, 2002.

[12] Winsberg, Elkin M., Macy J., Bordaz V., and Weymouth W., "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," *Radiology*, vol. 89, pp. 211–215, 1967.

[13] Marx C., Malich A., Facius M., Grebenstein U., Sauner D., Pfleiderer S.O.R., and Kaiser A.W, "Are unnecessary follow-up procedures induced by computer-aided diagnosis (CAD) in mammography? Comparison of Mammographic diagnosis with and without use of CAD", *European Journal of Radiology*, vol. 51, pp. 66–72, 2004.

[14] Sajda P., Spence C. and Pearson J., "Learning contextual relationships in mammograms using a hierarchical pyramid neural network," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 239-250, 2002.

[15] Breast Cancer Facts and Figure, http://www.cancer.org/research/cancerfactsfigures /breastcancerfactsfigures/breast-cancer-facts-and-figures-2011-2012,Date:01/03/2016

[16] Breen N., Gentleman J., Schiller J., "Update on mammography trends: comparisons of rates in 2000, 2005, and 2008," *Cancer*, vol.10, pp.2209-2218, 2011

[17] Lewin L.M., Orsi C.J., and Hendrick R.E, "Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer", *AJR Am J Roentgenol*, vol.3, pp.671-677, 2002.

[18] Pisano E., Gatsonis C., and Hendrick E., "Diagnostic performance of digital versus film mammography for breast-cancer screening", *The New England Journal of Medicine*, vol. 353,pp.1773-1783, 2005.

[19] Reston V.A. "Breast Imaging Reporting and Data system", 3rd edition.: *American College of Radiology*, 1992.

[20] Boyd N, Byng J., Long R., Fishell E, and Yaffe M., "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening study, *J Natl Cancer Inst* . vol. 87, pp. 670-675, 1995.

[21] Tao E., and Sklansky J., "Analysis of mammograms aided by database of images of calcifications and textures", *Proc SPIE* , pp.988-995, 1996.

[22] Hubbard R., Kerlikowske K., and Flowers C., "Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study", *Ann Intern Med*, vol.155, pp. 481-492, 2011

[23] Doi K., MacMahon H., Giger M., and Hoffmann K., "Computer-aided diagnosis and its potential impact on diagnostic radiology," *Computer-Aided Diagnosis in Medical imaging. Amsterdam, the Netherlands: Elsevier Science*, pp 11-20, 1999

[24] Aiello EJ, Buist DS, White E, and Porter PL," Association between mammographic breast density and breast cancer tumor characteristics", *Cancer Epidemiol Biomark Prev* vol.3, pp.662–668,2005

[25] Wolfe JN, Breast patterns as an index of risk for developing breast cancer. *Am J Roentgenol* , vol.6,pp.1130–1137, 1976

[26] Li H., Giger M.L., Huo Z., Olopade O.I., Lan L., Weber B.L., and Bonta I. , "Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location", *Med Phys* vol.31(3), pp.549–555, 2005

[27] Bovis K., and Singh S., "Classification of mammographic breast density using a combined classifier paradigm", *In 4th International Workshop on Digital Mammography*, pp.1–4, 2002

[28] Petroudi S, Kadir T, and Brady M ,"Automatic classification of mammographic parenchymal patterns: a statistical approach", *In: proceedings of the 25th annual international conference of the IEEE on engineering in medicine and biology society* vol.1, pp.798–801, 2003

[29] Hapfelmeier A, and Horsch A "Image feature evaluation in two new mammography CAD prototypes", *Int J CARS*, vol.6, pp.721-735, 201

[30] Oliver A, Freixenet J, Marti R, Pont J, Pérez E, Denton ER, and Zwiggelaar R, "A novel breast tissue density classification methodology", *Information technology in biomedicine*. IEEE Trans, vol.12(1): pp.55–65,2008

[31] Zacharaki E I, Wang S, Chawla S, Yoo D S, Wolf R, Mehem E R, and Davatzikos C, "Classification of brain tumor and grade using MRI texture in a Machine Learning technique", *Magn. Reson. Med.*, vol. 62, pp.1609-1618, 2009.

[32] Georgiardis P, Cavouras D, Kalatzis I, Daskalakis A, Kagadis G C, Malamas M, Nikifordis G, and Solomou E, "Improving brain tumor characterization on MRI by probabilistic neural network on non-linear transformation of textural features", *Comput Meth Prog bio.*, vol.89, pp.24-32, 2008.

[33] Ciatto S, Bernardi D, Calabrese M, Durando M, Gentilini MA, Mariscotti G, and Houssami N, "A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification" *The Breast*, vol.21(4), pp.503–506,2012

[34] Strickland R.N and Hahn H.I. "Wavelet transforms for detecting microcalcifications in mammograms." *IEEE Transactions on Medical Imaging,* 15:218-229, 1996.

[35] Zhang W, Yoshida H, Nishikawa RM, and Doi K., "Optimally weighted wavelet transform based on supervised training for detection of microcalcifications in digital mammograms." *Medical Physics,* vol.25, pp.949- 956, 1998.

[36] Qian W, Kallergi M., Clarke L.P., Li H. D., Venugopal P., Song D., and Clark R.A. "Tree structured wavelet transform segmentation of microcalcifications in digital mammography." *Medical Physics,* vol.22,pp.1247-1254, 1995.

[37] Gurcan M.N., Yardimci Y., Cetin A.E, and Ansari R.. "Detection of microcalcifications in mammograms using higher order statistics." *IEEE Signal Processing Letters,* vol.4, pp.213-216, 1997.

[38] Cands E. J. and Donoho D. L. "A key to higher dimensional intermittency?" *Philosophical Transactions of The Royal Society, The Computer Journal,* pp.2495-2509, 1999.

[39] Netsch T. and Peitgen H. O. "Scale-space signatures for the detection of clustered microcalcifications in digital mammograms." *IEEE Transactions on Medical Imaging,* vol.18, pp 774-786, 1999.

[40] Subhash J, Shankar K.R., and Karnan M, "Detection of Microcalcification in Mammograms using Soft Computing Techniques", *European Journal of Scientific Research,* vol. 86, pp.103-122, 2012

[41] Karssemeijer N, "Automate classification of parenchymal patterns in mammograms", Physics *in medicine and Biology*, vol.43, pp.365-389, 1998

[42] Mustra, Grgic M., and Delac K., "Breast density classification using multiple features selection", *Automatic, Journal of control measurement Electronics, Computing and communication*, vol. 53, pp. 362-372, 2012.

[43] Sharma V., Singh S.,"CFS SMO based classification of using multiple texture models", *Medical Biological Engineering and Computing*, vol.52, pp. 521-529, 2014

[44] Liu Q., Liu L., Tan Y., Wang J. and Ni H., "Mammogram density estimation using sub-region classification", In proceedings of 4th International conference on Biomedical engineering and Informatics (BMEI), pp. 356-359, October 2011

[45] Sampaio B., Moraes W., Silva C., and Gattass M., "Detection of masses in mammogram images using CNN, geostatistic functions and SVM," *Computers in Biology and Medicine*, vol.41, pp. 653–664, 2011

[46] Hall M.A., and Smith L.A., "Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper", *In proceedings of the twelfth international Florida artificial intelligence research society conference*, vol.235, pp.235–239, 1999

[47] Haralick R.M., Shanmugam K., and Dinstein I.H., "Textural features for image classification", *IEEE Trans Syst Man Cybern* , vol.3,pp.610–621, 1973

[48] Heath M., Bowyer K., Kopans D., Moore R. and Kegelmeyer P. J.," The digital database for screening mammography." *In proceeding. 5th int. workshop on digital mammography. Medical Physics Publishing*, pp. 212–218, London, 2001;

[49] Tuceryan M. and A.K. Jain, "Texture Analysis" from The Handbook of Pattern Recognition and Computer Vision, 2nd edition, (Editors C.H. Chen, L.F. Pau and P.S.P Wang), *World Scientific Publishing Co.*, pp. 207-248, 1998.

[50] Kadah Y. M., Farag A., Zurada J M, and Badawi, " Algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images", *IEEE Transactions on Medical Imaging*, vol.15(4), pp.466-478, 1996.

[51] Christodoulou C. I, Pattichis C. S, and Pantziar M., "Texture based classification of atherosclerotic carotid plaques", *IEEE Trans* carotid *Med Imaging*, vol.22, pp.902–12,2003.

[52] Stoitsis J, Valavanis I, Mougiakakou S G, Golemati S, and Nikita, " A aided diagnosis based on medical image", *Nuclear Inst Meth Phys Res*, vol. 569, pp.591–595, 2006

[53] Miroslav Benco, Robert Hudec , Patrik Kamencay , Martina Zachariasova, and Slavomir Matuskal , " An Advanced Approach to Extraction of Colour Texture Features base on GLCM", *International Journal of Advanced Robotic Systems*, 2014.

[54] Galloway M.M., "Texture analysis using gray level run lengths," *Comput. Graphics Image Process.* , vol.4, 172–179, 1975.

[55] Tang X., "Texture information in run-length matrices", *IEEE Trans Image Process*, vol 7, pp.1602–1609, 1998.

[56] Wu C.M., and Chen Y.C., "Statistical feature matrix for texture analysis" *Comput Vis Graph Image Process*, vol. 54(5) ,pp. 407–419, 1992

[57] Gabor D., "Theory of communication", *Journal of the Institution of Electrical Engineers – Part III: Radio and Communication Engineering*, vol. 93, pp.429–457, 1946.

[58] Manjunath B.S., and Ma W.Y., "Texture features for browsing and retrieval of image data", *IEEE Trans Pattern Anal Mach Intell*, vol.18, pp. 837–842, 1996.

[59] Gennari J., Langely P. and Fisher D., "Models of incremental concept formation,‖ Artifical Intelligence, vol.(40), pp. 11-61, 1989

[60] Support Vector Machines (SVM), http://in.mathworks.com/help/stats/support-vector-machines-svm.html?refresh=true, Date: 15 Dec 2015.

[61] Tadiou K.M., "Artificial Neural Networks "http://futurehumanevolution.com/artificial-"intelligence-future-humanevolution/ artificial-neural-networks, 2014.

[62] Sakka E., Prentza A. and Koutsouris D., "Classification algorithms for micro calcifications in mammograms" (Review), Oncology Reports, *Computational analysis and decision support systems in oncology*, vol.15, pp.969-1108, 2006

[63] Shimizu A., Hasegawa J., and Toriwaki J., "Minimum directional difference filter for extraction of circumscribed shadow in chest x-ray images and its characteristics, *Trans of IEICE*, vol. J76D-II, pp.241-249 ,1993