*IOSR Journal of Dental and Medical Sciences (IOSR-JDMS)*
*e-ISSN: 2279-0853, p-ISSN: 2279-0861.Volume 23, Issue 6 Ser. 3 (June. 2024), PP 21-51*
*www.iosrjournals.org*

# Approaching Of The Machine Vision In The Classification Of Breast Cancer Using Texture Analysis

Dr Khalid Mehmood/ Prof Dr Li Zhen Nan
*Doctor In Surgery In*
*Northern Jiangsu People Hospital ,Yangzhou city China*

***Abstract***
*The main objective of this research is to develop a diagnostic system, based on standard or objective parameters rather than non-standard or subjective parameters, which are already being employed by radiologists, for the classification of abnormalities present in mammograms, as benign or malignant. Radiologists differentiate biological behavior of these abnormalities on the basis of visual parameters such as size, shape and boundaries of the tumors. A benign tumor has small size, well-defined margins and homogenous texture,whereas, a malignant tumor typically has larger size, poorly margined and heterogeneous texture. Due to the limitations of human perception all these parameters become subjective, which cause a high risk of misinterpretation, inter and intra- observer variation, for correct decision. Considering all these factors, development of a compact system is required; (i) to accurately classify malignant and benign abnormalities within a reasonable time and cost,*
*(ii) to increase diagnostic consistency by providing an objective (rather than subjective) evaluation. To meet the objectives a CAD system has been developed and employed ontwo types of data; (a) test data and (b) experimental data.*
*In test data, five wheat varieties are differentiated by this system. For this purpose five types of 77 statistical textural features, which may be grouped as; first order (histogram) features, second order (GLCM) features, higher order (GLRM) features, autoregressive features, and gradient matrix based features are calculated from ROIs (8x8) (16x16) (32x32), and (64x64) under, μ±3σ and 1-99% normalization conditions by using MaZda software. The most relevant features for each size of ROI are selected by three approaches; Fisher's Co-efficient, Probability of Error plus Average Correlation Co-efficient, andMutual Information Co-efficient. In this way the most relevant 10 features are selected by each method. We received very poor results when data analysis capability is verified on the basis of 10 features selected by each method for each size of ROI except (64x64), by three multivariate techniques; PCA, LDA, and NDA under both normalization conditions, by a software „B11", integrated with MaZda.To improve the results, a set of 19 features is obtained by merging the features selected by each approach. An excellent clustering result with an accuracy of 99.67% is received, when data of these 19 features extracted from ROI (64x64) under 1-99% normalization, is deployed to NDA projection space. By using supervised classification approach, artificial neural network (ANN) the system is trained and tested on the basis of 70% and 30% of input data respectively. We received an accuracy of 99.90% and 93.11% in training and testing phase respectively.*
*On the basis of results for test data analysis, it is concluded that the proposed CAD system produces the best result for large ROI window size when a combined set of features is deployed in NDA projection space. The experimental data (mammograms) is analyzed under these settings.*
*The mammography data is consisted of two types of images, Craniocaudel (CC) and Medioletral Oblique (MLO) view images. Unlike to other researchers in this work both types of images are considered in separate sections.*
*As the experimental data (mammograms) has fine and micro-texture, so, initially abnormal regions in CC view images, marked by radiologists, are tried to be analyzed on the basis of combined set of features (discussed above for the case of test data), extracted from ROI (8x8) under above mentioned both normalization conditions. As NDA approach based on ANN classifier and a number of options are available on „B11" software to configure this classifier. Data analysis capability of selected features under different architectural settings of ANN on the basis number of neurons in input hidden layer and learning rate „η" in NDA projection space is tried. Then the classifier is trained and tested on the basis of architectural settings for which the best clustering is received, by splitting data in 70/30 ratio respectively.*
*For ROI (8x8) we received testing accuracy of 91.18% when the classifier is configured with 2 neurons in hidden layer and learning rate is set at 0.15 when the features are extracted under μ±3σ normalization condition. For same of size of ROI when features are extracted under 1-99% conditions, the best testing accuracy of 88.44% is obtained for same architectural settings (2 neurons and η=0.15).*
*As the performance of the system for ROI (8x8) is not satisfactory, so, we tried to classify same data by*

*extracting features from ROIs (16x16) under the both normalization approaches. Following the same procedural steps we received an accuracy of 92.56% for testing phase, when the classifier is configured with 2 neurons in hidden layer at learning rate 0.20 and the features are extracted under 1-99% normalization. We received excellent classifier testing result with an accuracy of 97.55% when the classifier is configured with 3 neurons in hidden layer at learning rate 0.15 and the features are extracted by applying μ±3σ approach.*
*For MLO view images we obtained a testing accuracy of 84.41%, when the classifier is configured with 5 neurons in hidden layer with learning rate is set at 0.22 and the data is normalized by μ±3σ approach.*

---

Date of Submission: 28-05-2024 Date of Acceptance: 08-06-2024

---

**Dissertation Overview**

This thesis is comprises of six chapters. The brief description of these chapters is presentedbelow:

Statistics regarding to breast cancer are presented in initial part of Chapter 1, which highlight the importance of this work. Statistics are also presented which show that early detection of disease is the only way to reduce the mortality rate, and mammography is the most reliable modality, which is being used for this purpose. Problems and motivations are discussed in the next section, after this research objectives and scope of this research are presented. In the last section an outline of proposed frame work, which based on objective parameters rather than subjective, for the classification of breast cancer in digital mammograms has been discussed. Finally, at the end of this chapter a flow chart ispresented to highlight the main steps of proposed CAD system.

In Chapter 2, first of all anatomy of female breast is discussed, after this microclacifications (MCs) are discussed, which are considered as the most initial symptomof breast cancer. In the next sections cancer types, stages regarding to the treatment, and anintroduction of breast grading system (BIRADS) is presented. In the next section X-rays based breast imaging modality, mammography, along with different findings of mammograms, is presented. Then, a brief description of other imaging modalities, being used for this purpose, is presented. At the end of this chapter CAD system for the detectionand differentiation of breast cancer along with a brief literature survey regarding to the application of this system is produced.

As this work is mainly based on texture analysis, so, in the start of Chapter 3 we have discussed the different issues regarding to texture definition and its elements. After this types of texture and analysis techniques are presented. As for this work, we are mainly concerned to statistical approach so, a detail description of GLCM, GLRM, GLDM, along with their features is presented there. In the last of this chapter, a brief description regarding interaction of human perception with textures is discussed.

In Chapter 4 methodology of this research and the used software MaZda is introduced. In start of this chapter importance of CAD system and an overview of implemented algorithmis presented. After this a brief introduction of procedural steps employed in this experimental work like development for ROIs, images normalization, statistical textural features extraction, features selection, features reduction, and ANN classifier along with their theoretical background is provided. In the next section, the designed CAD system is utilized to differentiate five wheat classes (as test data). The main objective of this implementation was to verify the performance of the proposed system, and, we received the best results with an accuracy of 99.90% and 93.11% for training and testing respectively, when features were extracted from ROI (64x64), the most relevant features were selected by a hybrid model of features selection and NDA analysis approach was employed.

In Chapter 5, detail of implementation of the proposed system on experimental data (digital mammograms) is presented. This detail is divided in three modules; in first moduledetail of preprocessing of data like marking of suspicious regions by radiologists, changeof image format, and development of RIOs etc., is presented. The second module has the detail of features extraction and features selection. First part of this module contains the detail of data analysis/ classification of Craniocaudel (CC) view images under two normalization approaches, μ±3σ and 1-99%, for the ROIs having window sizes (8x8) and (16x16), and second part contains analysis detail of Medioletral Oblique (MLO) view images. In third module, training and testing results under different architectural settings ofANN classifier are presented. For CC view images we received the best results with an accuracy of 99.21% and 97.55% for training and testing the classifier respectively, when the features were calculated from ROI (16x16) under μ±3σ normalization and for MLO these results were 91.59% and 84.41% respectively.

In Chapter 6 all the results of test data and experimental data are compared and discussed. At the end of this chapter a brief conclusion of the whole research and suggestions for future work are presented.

## I. Introduction

Breast cancer is considered as the most common breast disease and a leading cause of death among the women.

―*Cancer is a group of many related diseases that begins in cells. Normallycells grow and divide to produce more cells only when the body needs. This orderly process helps the body to keep healthy. Sometimes, however, cells keep dividing when new cells are not needed there. These extra cells form a mass or a tumor. A tumor may be benign or malignant. Benign tumors do not spread to the other parts of body and are not life threatening. Malignanttumors can invade and damage nearby tissue, can spread other parts of body, a process called metastasis*‖[1].

The sensitivity of this problem has been highlighted by a number of researchers working inthis area. A brief review of literature, regarding to disease statistics, presented in the following section reflects the importance of this issue.

**Disease Statistics**

According to a recent survey in the USA a woman has 12.15% lifetime risk of being diagnosed with the breast cancer whereas in 1970s it was only 10%[2]. Statistics published by American Cancer Society revealed that one out of eight women, and one in twelve men develop breast cancer during their life. Between 1973 and 1999, incidence rates of this cancer were increased nearly 40%, however, between the period of 1989 to 1995 mortality rates decreased by 1.4% per year and 3.2% afterwards, due to early detection of the disease[3]. In the year 2009, approximately 192,370 new cases of invasive breast cancer and 62,280 situ breast cancers were reported. Moreover, 40,170 women were estimated to die with this cancer[4]. In 2011, approximately 200,000 women diagnosed with breast cancer, and 40,000 women were died with this disease during the year[5].

In Europe during 2006, breast cancer was reported as the most common cancer amongst female, with 429,900 cases (28.9% of all types of cancers) [6]. According to a survey report made on 100,000 women from 1995 to 1998 in the European Union, indicated that for this cancer approximately 39 deaths per year occurred and 40 deaths happened during 1985 to 1989, regardless of age, showing a reduction of mortality rate of 2.1%. This favorable trend is due to advanced screening and detection of disease at early stage [7].

The incident rate for this type cancer in Asian and African countries is significantly lower as compare to Europe and the USA. The lowest rate is in central Africa, where 16.5 women per 100,000 according to age standardized rate (ASR) [8]. In Saudi Arabia a total number of 10,214 cases of breast cancer were registered in 2005 [9]. In Malaysia, from 2003 to 2005 a total number of 11,952 new cases of this cancer were registered byNational Cancer Registry (NCR), and it is 18% of all new cases of cancer. In women this cancer is accounted for 31.3% of total number of all types of cancer [10].

In Pakistan it is estimated that one out of nine women is likely to be diagnosed as breast cancer patient, which is the highest rate of incidence in Asia, by ASR it is 69.1 cases per 100,000 women for the period 1998 to 2002. According to Karachi Cancer Registry report breast cancer accounts for approximately 34.6% of all cancer cases in the city making it themost common cancer in women [11]. The disease is associated with significant mortality in Pakistan as the most cases are diagnosed at advanced stage, where the treatment is limited. It happens due to the lack of knowledge regarding to risk factors, awareness, and screening facilities [12]. Fig. 1.1 presents an international comparison for age standardized incidences of this cancer per 100,000 women during 2007-2008, which shows that it has maximum rate in the USA which is 92.1, while the minimum rate of 20.8 at Korea.

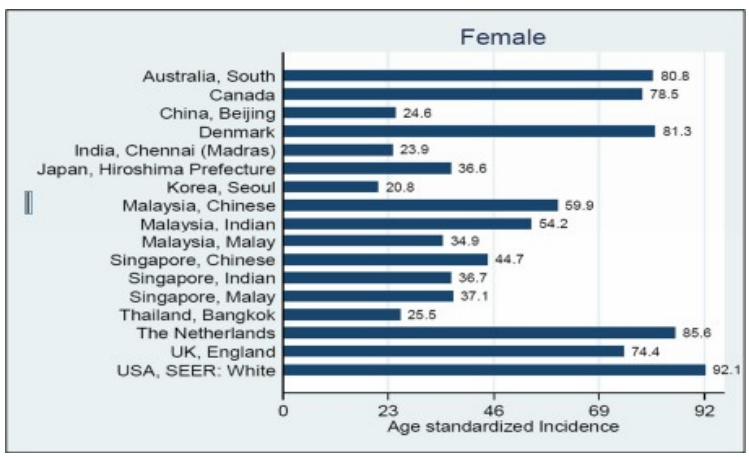



**Fig. 1.01: International comparisons--Age Standardized breast cancer incidences per100,000 population [10].**

Up till, the basic factors, causing this cancer are unknown, however, some of the well- recognized risk factors are postmenopausal obesity, use of combined estrogen and progestin menopausal hormones, alcohol consumption, and physical inactivity [13]. Since the causes of this disease are still a mystery, hence, its primary preventions are seemed to be impossible, however, early detection is the key to improving the breast cancer prognosis[14-15].

Early detection of disease leads to significant improvements in conservative treatment and its control. Breast carcinoma is usually curable by good surgery when primary lesion is small and only one or two auxiliaries are involved, but when auxiliary metastases of considerable extent are present, the rate of cure is greatly diminished and cure becomes impossible when distant metastases have developed. For the patients in the last category improvements even by surgical techniques offer no hope, whatever, because the disease is far beyond the surgeon‟s approach.

Breast cancer survival rates decrease about 95% with the lesion size of 0.5cm to only 75% or less when the cancer is treated at size of 2.5cm, unfortunately the cancers are detected by patients themselves when the size of cancer lesion is more than this size [16]. Due to this reason, Haggises advises a regular breast screening at least once in a period of three months to detect carcinoma at early stage [17].

To detect the cancer at early stage a number of imaging modalities and techniques are being used for breast screening. X-rays mammography and clinical breast examination (CBE) are considered the best methods for breast screening. Ultrasound and MRI are also being used as secondary tools to have further details of suspicious findings from these mammograms. Some other non-optical techniques like positron emission tomography (PET), electrical impedance tomography (EIT), and thermal imaging are also being used in this regard [18]. More recently, Digital Mammography, where image is acquired in digital form on a screen rather than a photographic plate, is considered as the most reliable tool for this purpose [19].

Literature survey reveals that during the recent decades the incident rate of this cancer has increased, but there is a decline in mortality rate among the women of all ages. This trend of reduction in mortality is associated with adoption of advanced mammography screening techniques, by which sign of breast carcinoma, like microcalcifications (MCs) at the earliest stage may be detected [20-21]. The presence of small mineral deposits in the breast tissue which usually appear in the form of groups, commonly referred to as microcalcifications, can constitute a first sign of breast cancer [22]. Microcalcifications may be the only mammographic sign of non-palpable breast cancer, which appears as bright spots in an image. Another symptom of breast cancer which is detectable by mammograms is the mass lesions. These are the two most (MCs and lesions) common breast cancer indicators, which are present in 30% to 50% of all cancer found mammographically [23].

**Problems and Motivation**

No doubt mammography is one of the most reliable technique for the detection of breast carcinoma, however, detection of MCs (which is the only mammographic sign of breast cancer), in mammograms by naked eye is a difficult task due to, small size in the range of `0.05mm to 1.0mm [24], low contrast, different shapes and subtle nature [21]. Small MCs ranging from 0.1mm to 0.2 mm can hardly be seen in mammograms because of noise and their superimposition on the breast parenchymal texture. Low contrast MCs are often overseen, because, a region which differs in luminance from its surroundings by less than 2% is indistinguishable to human eye [25]. MCs often appear in an inhomogeneous background described by the breast tissue. Some parts of the background may have such features that they appear brighter than MCs in fatty tissues [26].

Detection of lesions is also a difficult task, because these are hidden and present in dense glandular area of the breast tissue. Sometimes it is difficult due to variation in shape, size and dimensions. These are also indistinguishable due to their features (heuristics), can be obscure as they are similar to the normal inhomogeneous tissue [27].

Because of above discussed reasons 10% to 30% of lesions are misinterpreted during screening procedures of mammograms. According to another study report, interpretation error (e.g., the radiologist sees the cancer, but thinks it is benign) accounts for 54% of missed cancers [28]. It reflects that even mammographically it is very difficult to differentiate benign lesion from malignant ones.

For this reason, it is difficult for radiologists to provide accurate and uniform evaluation for the enormous number of mammograms. Due to some limitations of human observers: 10-30% lesions are overseen during routine screening [29]. Missed cases are increased to 50% if minimal signs are taken into account [30].

It is also considerable that sensitivity and specificity of mammography range between 70%to 96% and 90% to 95% respectively. These factors are highly related to technical quality aspects and specially, to radiologist‟s detection capacity, which depends on their awarenessand experience. It has been proved by several clinical studies that between 30% to 70% of cancer diagnosed in a screening program are visible in retrospect mammograms previously, interpreted as normal and almost all these cases are due to detection error [14, 31].

Hence, to identify the cancerous tissue in breast, usually a biopsy is performed. In this procedure a suspicious tissue is removed from breast by surgical operation for examination. By pathological examination, the abnormality in that tissue is decided, as malignant (cancerous) or benign (not cancerous) tissue. Clinical studies reveal that positiveproductive value (PPV), which is the ratio of total cancer found to the total number of biopsies are 15% to 30% [32]. Between 2 and 10 women are who can lead to malignant lesions not being biopsied and benign lesions biopsied [33]. Moreover, surgical, needle core and fine needle biopsies are expensive, invasive, time consuming and traumatic forthe patient biopsied. This low specificity results in relatively large inter-observer variability [34].

A radiologist differentiates biological behavior of breast tumors on the basis of visual parameters such as size, shape and boundaries of the tumors. A benign tumor has small size, well defined margins and homogenous texture, while a malignant tumor typically has larger size, poorly margined and heterogeneous texture. Due to the limitations of human perception all these parameters become subjective, which cause a high risk ofmisinterpretation and inter and intra- observer variation for correct decision [35].

Due to all these factors interpretation error rate of 30% is reported among the radiologists [36]. Analysis consistency may be achieved by the use of double reading of two or more than two radiologists, but it also increases the time and cost of mammography screening process [19, 37]. Analysis consistency begins to suffer from a subjective component of data evaluation that depends on factors, such as radiologist‟s training, experience and fatigue.

Considering the traumatic nature, cost of biopsy and relatively difficult task for radiologists to interpret mammographic images, it is desirable to develop a compact system; (i) to accurately classify malignant and benign abnormalities within a reasonable time and cost, (ii) to increase diagnostic consistency by providing an objective (rather than subjective) evaluation.

Computers have the ability to detect complex patterns and handle statistics in a large dataset, which are beyond the human perception. Computers also do not face the biologicallimitations like fatigue that a human would face by analyzing a large dataset images. Furthermore, a number of computers can be used in parallel mode to reduce the time needed to obtain a result. Diagnostic information from the medical images can be improved by designing computer processing algorithms, applications and developing software intelligently [38]. Algorithms can be developed to have consistent and objective result. Using machine (computer) vision approaches radiologists can incorporate computer output into their results. Literature survey reveals that this approach improves the radiologists‟ ability to differentiate malignant and benign abnormalities [14, 39-40].

Machine vision approach is based on the utilization of computer for texture analysis. Texture analysis is the combination of image processing and machine learning techniques. Image processing is widely being implemented to suppress the artifacts, eliminate noise, enhance the image and to extract statistical parameters for detection/classification of lesions and MCs in mammograms [41]. Reports are available in literature that radiologists perform better on enhanced images [21]. As the conventional enhancement methods are not quite precise, because they affect all the image changes, not only small light details like microcalcifications, so, by digitizing mammographic films and applying digital image processing algorithms, significant improvements in image interpretation are possible [42- 43]. Machine learning is being implemented for pattern classification purpose. Here a number of quantitative features are calculated from region of interests (RIOs), and then a set of the most relevant features is selected by using predefined statistical techniques. A classifier is trained and tested on the basis of these selected features for the knownabnormalities (lesions and MCs) present in mammograms.

**Research Objectives and Scope**

The aim of this work is to increase the diagnostic accuracy of radiologists for the detection and classification of abnormalities, present in mammographic images, as benign or malignant with the implementation of texture analysis techniques. Here the second available potential advantage of digital mammography, that is, quantitative analysis, rather than qualitative analysis, will be worked out. Currently available methods are time consuming, model based and depend upon non-standard qualitative parameters like, size, shape, contrast and homogeneity of the tumor. Because all these parameters based on visual assessment of radiologist, and human visual capability to differentiate within a wide range of textures is very limited [44-45], so, classification of tumors as benign or malignant suffers from inter- and intra-observer variations.

The general purpose of this work is to provide a computer assisted detection/classification method for radiologists, which is model independent, speedy and based on standard quantitative parameters to have more consistent results. The approach of this research is to apply statistical texture analysis methods to analyze mammogram datasets. The results of this approach are compared with manual analysis, performed by radiologists.

The specific aims of this dissertation are:
(a) Provide an automatic or semi-automatic application to radiologists for quantitative and quick analysis of mammograms.
(b) Quantify inter- and intra- observer variations, which occur due to visual assessment, and compare with this investigated approach.
(c) Explore the performance of this application in different automatic diagnostic regions.

The research presented in this dissertation aims to provide a computer based objective approach, by which radiologists and other clinicians can classify and differentiate abnormalities present in medical images (specially in mammograms) more accurately. This approach is based on the classification of gray level statistical texture features calculated on pixel level. This approach is advantageous because it does not depend upon contrast enhancement parameter model. This pixel to pixel approach (rather than region based approach) offers the potential for a more precise classification of tumors. The objective of this work is not to replace the human observers, but rather give them a tool to increase the accuracy. In addition this application may increase analysis throughput. Faster analysis may help to reduce delays in patient receiving therapy. The use of computer based analysis may also be used for efficient evaluation of the potentially hundreds or thousands of datasets involved with clinical studies.

In this dissertation we present a machine vision approach to use pixel based classification of abnormalities in mammograms which does not rely on visual parameters like size, shape, intensity and boundary conditions of a lesion. Fig. 1.2 illustrates an overview of the proposed approach.

**Overview of Frame-Work**

**Experimental Data**

The inputs of this proposed experimental work are raw (unprocessed) digital mammograms and machine vision is applied for the classification of malignant and benign abnormalities. The required data (digital mammograms) is obtained from the Al-Noor Diagnosing Center Shadman, Lahore, Pakistan. Total 600 patient‟s data is obtained from the said center. The data of each patient is consisting of four digital images; two (left and right) images of mediolateral oblique view and two (left and right) craniocaudal view.

**Regions of Interest (RIOs)**

Each mammogram has a number of anatomical parts of the breast, which are visually differentiable due to intensity variation. Denser tissues like muscles, fibroglandular tissues, vescular tissues, malignant and benign masses are displayed brighter, whereas, areas containing skin and fat appear darker in mammograms. This work is mainly concerned to the regions which contain some type of abnormalities (malignant and benign masses), which are phrased as regions of interest (RIOs). These RIOs are marked/identified with the help of expert radiologists of Department of Radiology and Diagnostic Imaging of Bahawal Victoria Hospital, Bahawalpur, Pakistan.

**Features Extraction**

In this research work classification of breast abnormalities (lesions and MCs) is performed on pixel based information which does not rely on any predefined model. These pixel related information define the texture of the ROIs in the form of intensity variation in that region. Digital mammograms are known to contain important textural information. In machine vision and image analysis these texture information are described by quantitative measures called statistical texture features or parameters.

The performance of machine vision/learning algorithms is highly correlated to these implemented quantitative measures [46-48]. Haralick [49] texture parameters are used in his work (which are the best suited for the analysis of fine and micro-textures like biomedical images) to find textural information of the ROIs. The said Haralick‟s parameters are extracted from the computed Spatial Gray Level Co-occurrence Matrices (GLCM).

**Features Selection and Reduction**

In the result of above-mentioned computational procedure (Section 1.4.3) a huge number of statistical textural parameters are generated from the describe ROIs. Statistically all the computed features are not equally important for texture analysis, so, it is necessary to find out the most relevant set of features. When dealing with many features, manual selection of the most relevant features is very difficult, so, it needs some automatic features selection methods. In literature a number of feature selection methods are available, which are being used to eliminate irrelevant or redundant features. Its choice depends upon various dataset characteristics: (i) data type (ii) data size, and (iii) data quality [50]. Based on different criteria from these characteristics, following three feature selection methods are adopted in this work; (a) *Fisher Co-efficient (F),*

(b)*Probability of Error and Average Correlation Co-efficient (POE+ACC),* and (c) *Mutual Information (MI),* to select the most relevant features. MaZda [51] software is used to calculate and select the said features.

With the help of B11 (an application integrated with MaZda software for statistical analysis) the following three feature vector transformations are applied to transform the data to a new feature space, in order either to further reduce the feature dimensionality or to increase the discriminative power of the features. These are *Principal ComponentAnalysis (PCA), Linear Discriminant Analysis (LDA),* and *Non-Linear Discriminant Analysis (NDA).*

**Pattern Classification**

Once a set of parameters (features) is selected, a pattern classification algorithm is employed to classify the tissues in mammograms. The goal of pattern classification is to assign every pixel in the image to a class. In this work there are two classes normal (benign) and pathological (malignant). The classification system must go through learning process where the system is presented with sets of features (parameters) and their associated classes. Although a number of pattern recognition algorithms exit, in this study we used artificial neural network *(ANN)* as tissue classifier. The output of *ANN* classifier isa probability map that provides the estimated probability that an imaged tissue in malignant. Given this probability map, output images may be produced to convey a human readable result that a radiologist may use for his interpretation.

**Significance of Work**

In digital mammograms it is observed that most of the lesions and MCs are stellate in appearance, their masses are irregular with ill-defined boundaries, and their size can vary from a few millimeters to several centimeters in diameter. All these factors and the reasons discussed in Section 1.2 make the detection and classification of lesions as malignant and benign a challenging task.

Applying machine vision approach for the classification of abnormalities in mammograms requires calculation of features from the suspicious spots (ROIs). In general, it is difficultto determine the size of ROI that should be used to calculate the said textural features, because, if a large ROI is used, and size of mass/ MC is small, then irrelevant portion of the organ may affect the results, similarly, if ROI is very small for large masses/MCs, it may cause missing of relevant information. Thus, the first contribution of this work is:
To find out the most suitable size of ROI form the lesions/MCs for the purpose of textural features calculation.

Identification of lesions and MCs in mammograms itself is a challenging task for a radiologist due to variability in their size, shape, and dimensions. Thus, the second contribution of this work is:

To demonstrate the advanced machine learning techniques like problems. Artificial neural networks *(ANNs)* can effectively solve the pattern recognition and classification.

As an ANN may be configured on the basis of architectural parameters:
To find out the best architectural settings of *ANN* to have the best results.

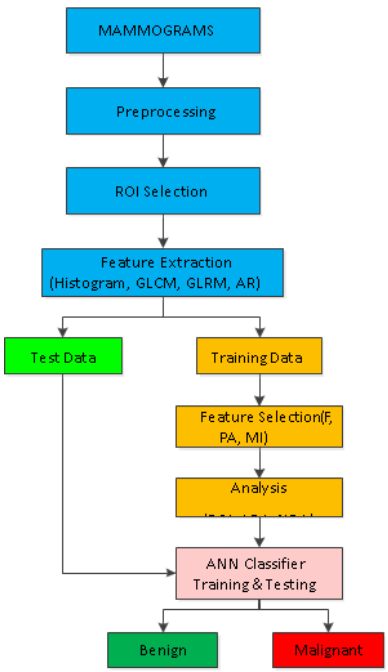



**Fig.1. 2: Flow chart of purposed experimental work**

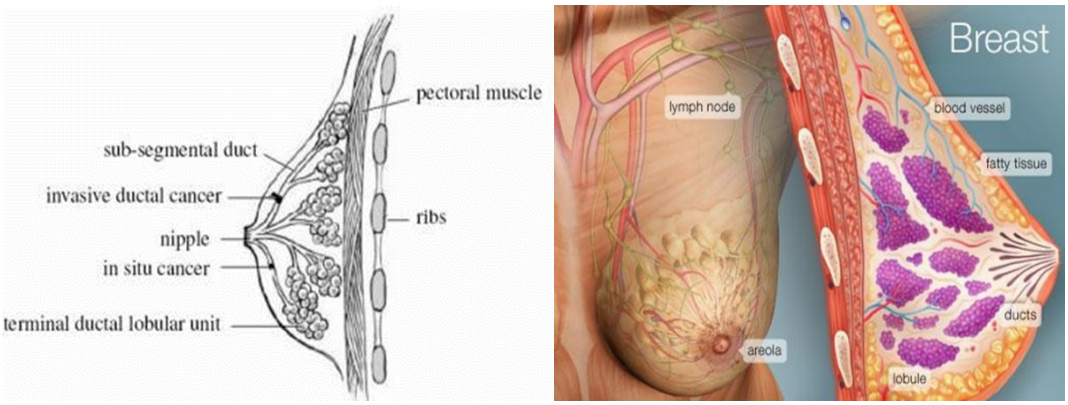
## II. Breast Cancer And Imaging Modalities

**Breast Anatomy**

The breast is a mound of glandular, fatty, and fibrous tissues located over the pectoralis muscles of the chest wall and attached to these muscles by fibrous strands. A layer of fat surrounds the breast glands and extends throughout the beast. The actual breast is composed of fat, glands with capacity for milk production, blood vessels, and milk ducts totransfer milk from the glands to the nipple [52].

It consists of two the most important components, first is related to milk production, whichis known as epithelial component and this component is composed of fat and connective tissue, which protects and supports the structure of the breast [53].

The structure of epithelial component is like a tree and all its branches terminate at nipple. Lobules may be considered as the leaves of this epithelial tree, which are secretory units ofthe breast. Each lobule has a number of acini connecting to an intra-lobular duct. The acini have two types of cells which are termed as epithelial and myo-epithelial.




**(a)** **(b)**

**Fig. 2.1:Breast Anatomy (a),(b) [54]**

The epithelial cells secrete a variety of glycol-proteins and during milk they also produce lactation. Each intra-lobular duct connects to an extra-lobular duct and this together with the lobule, which is termed as terminal ductal lobular unit. The extra-lobular ducts join together and form sub-segmental ducts, which finally form segmental ducts, and these drain milk from different lobes. A breast has 15 to 20 lobes having pyramid shape with apex towards nipple [55].

The non-epithelial components consist of fat tissue mainly, there are no muscles in a breastbut there are a number of muscles under and underneath the breast. These muscles work together a ligament, which is called cooper ligament. This cooper ligament supports the weight of the breast [55].

Breasts have lymph vessels, which are important to protect the body from the diseases. It isa clear fluid which contains tissue fluid waste products and also immune system cells. It consists of lymph nodes and lymph vessels which carry lymph to lymph nodes. Most of thelymph vessels that pass through the breast and transport lymph to the lymph underneath the arm pit, are called axillary nodes and which are in the chest are termed as mammary nodes. The other lymph nodes which are above and below collarbone are called supraclavicular or infraclavicular nodes respectively. Lymph veins may also transport possible disease to lymph nodes, which might increase the spread of the disease, such as breast cancer [56].

As discussed in Section 1.1 of this thesis, breast cancer is the most common type of cancer throughout the world and a major cause of death among the women. This is a non-preventive disease because up to now cause of this disease is a mystery. The best strategyto handle this situation and reduce the death rates is the early detection of this disease. Mammography is the most effective modality which is currently available to detect tumors in breast tissue, which can indicate potential clinical problems, such as breast architectural distortion, benign densities, and microcalcifications (MCs) *etc*. So far, there are the two most common features which are typically related to the breast tumors are MCs and mass lesions, which we have discussed in the following sections.

**Microcalcifications (MCs)**

Two types of abnormalities are observed during breast screening; one is the localized dense tissue which is usually termed as mass, second is the bright spots which are called microcalcifications [57]. Minerals are deposited in breast tissue in the form of small clusters are called microcalcifications. These microcalcifications are produced by the cells due to some benign or malignant process [58].

"*The radiological definition of clustered microcalcification is the presence of more than tree microcalcifications in 1cm² area*‖ [59].

The presence of these microcalcification clusters in mammograms is one of the symptoms of breast cancer at early stage [60]. Because these have higher X-rays attenuation as compared to normal breast tissue, therefore, MCs appear as a small, localized granular groups of bright spots in mammograms [61]. Detection of MCs in mammograms is an easy task, but it is very difficult to distinguish between malignant and benign MCs clusters due to fuzzy nature, low contrast and low differentiability form their surroundings [29].

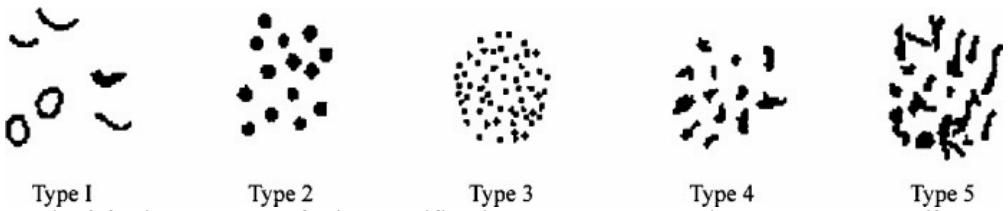
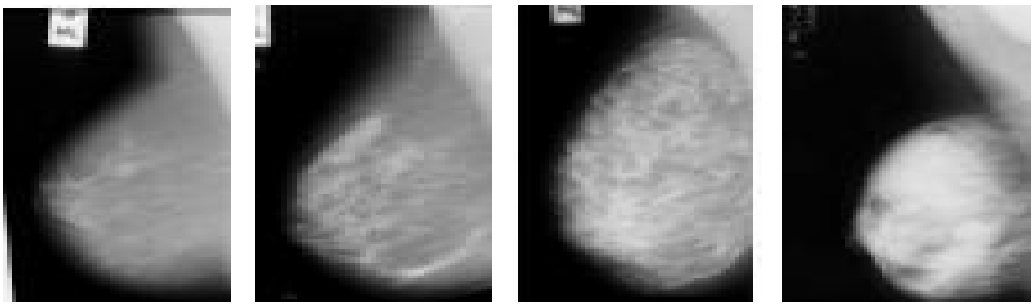



**Fig. 2.2:Five patterns of microcalcification clusters present in mammograms [62].**

The probability of MCs cluster to be a malignant one is high if the cluster has a high degree of polymorphic and branching nature. Generally, such type of MCs arise in ductal carcinoma. MCs could be found in ducts, alongside ducts, in the lobular acini, in vascular structures, in the interlobular stroma, in fat, or in skin. Depending on their etiology and location they could be punctuate, branching, linear, spherical, fine, coarse, cylindrical, smooth, jagged, regular in size, in shape or heterogeneous.

Since last 30 years the characteristics of MCs are being studied, these efforts led to establishing a system named; Breast Imaging Reporting and Data System (BIRADS) [34]. Now American College of Radiolog‟s Breast Imaging Reporting and Data System (ACR/BIRADS) is considered as a standard for the assessment of mammograms [63]. This system divides the interpretation of mammogram images in four categories according to densities as under:

ACR/BIRADS I: for almost entirely fatty breast (1-25%).
ACR/BIRADS II: for some fibro-glandular tissue (26-50%).
ACR/BIRADS III: for the breast which heterogeneously dense (51-75%).
ACR/BIRADS IV: for extremely dense breast (76-100%).



**(a) (b) (c) (d)**

**Fig. 2.3:Example of Image illustrating BIRAID stages (a) stage I, (b) stage II, (c) stage III, stage IV.**

Along with the characteristics, discussed above there are also some image related issues which make interpretation of MCs a tedious job for a radiologist:

MCs have very small size in the range of 0.1mm to 1mm, with the average 0.3mm approximately. Some isolated MCs having a size smaller than 0.1mm are indistinguishable due to high frequency noise [29].

MCs have various sizes, shapes, and distribution patterns [52].

MCs may have low contrast, due to which it becomes very difficult to distinguish suspicious spot from its surroundings [21, 29].

Suspicious spots are invisible due to overlapping of some dense tissues, and/or thickening of skin, especially in younger women. The dense tissues like fibrous strands, breast borders, and hyperprophied lobules that look like microcalcificationsmay easily be misinterpreted [29].

MCs may be closely connected to surrounding tissues which may not be segmentedwith simple methods [29].

Due to all these factors a radiologist cannot properly differentiate MCs as benign and malignant, which leads to the practice of un-necessary biopsies. It is the reason that only 20-30% biopsies out of all recommended cases turn out as malignant nature [64].

Hence, it is of crucial importance to develop a system by which these MCs may be classified as malignant or benign.

**Categories of Breast Cancer**

Every breast has 15 to 20 sections, which are called lobes. These lubes are further having smaller sections, named as lobules. The both lobes and lobules are connected to each other by thin tubes, called ducts. The most common type of breast cancer is the ductal cancer (the cancer related to these ducts). If the disease origin is in lobes or lobules then it is called lobular cancer.

Breast cancer is categorized as; (a) non-invasive (in situ or localized), (b) invasive (infiltrating).

Non-invasive cancer is confined to the site of origin. It includes the ductal carcinoma in situ (DCIS); sometimes it is termed as intra-ductal carcinoma and lobular cancer in situ (LCIS). The term situ refers to the disease condition that it has not spread to the other parts of the body. In DCIS the cancer cells are in the lining of the duct. It is a non-invasive cancer, if it is not treated properly at its early stage, it may progress to invasive cancer. Although LCIS is a non-invasive cancer but it is an indicative of the risk of invasive cancer in both breasts [65].

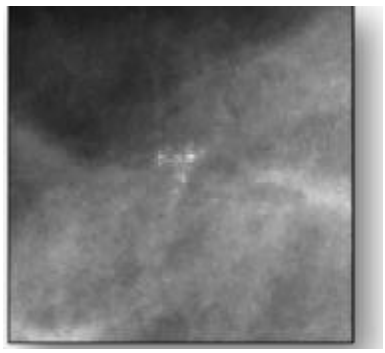


**Fig. 2.4: Example of DCIS**

Invasive breast cancer has the tendency to spread (metastasize) beyond the basement membrane. This membrane covers the underlying connective tissue in thebreast. This tissue contains a number of blood vessels lymphatic channels which are main source to carry the cancer cells beyond the breast. This type of cancer includes infiltrating ductal carcinoma and infiltrating lobular carcinoma. Infiltratingductal carcinoma is the type of breast cancer which penetrates the walls of duct. It comprises 70% to 80% of all types of breast cancer. Infiltrating lobular carcinoma spread though the wall of a lobule. It accounts for 10% to 15% of all types of breastcancer. Sometimes it can appear both in breast and in several locations [65].

**Stages of Breast Cancer**

On the basis of tumor size American Joint Committee on Cancer has defined TNM for the staging of breast cancer. Parameter „T‟ is related to the size of tumor, either the tumor hasspread or not to the axillary lymph nodes is specified by „N‟ parameter, and whether or not the tumor has metastasized is specified by parameter „M‟ (that is the tumor spread to moredistant parts of the body) [66].

For the purpose of treatment, Union International Cancer Centre (UICC), has divided thebreast cancer in five stages as under:

Stage 0 (zero)-- this stage includes the cancer cells which are non-invasive, but these cells pose the long term risk of becoming invasive. Lobular carcinoma in situ (LCIS) and ductal carcinoma in situ (DCIS) belong to this stage.

Stage I (one) -- it is the early stage of cancer where the tumor is less than two cm and has spread to the lymph nodes under the arm; or the size of the tumor is 2cm to 5cm with or without spreading to lymph nodes under the arm; or the size of the tumor is more than the 5cm and it has not spread outside the breast.

Stage III (three) – it is a locally advance cancer in which tumor is greater than 5cm and has spread to the lymph nodes under the arm; or the cancer is extensive in the under arm lymph nodes; or it has spread to lymph nodes near the breast bone and other parts of the breast.

Stage IV (four) – this is the stage where the cancers are spread out of breast to the other organs of the body. In this stage the cancers are metastasized.

The above discussed disease stages show that seriousness of the disease increases, andsurvival rates decreases with the stages.

**Breast Imaging**

In the recent years the occurrence of breast cancer has increased but, there is a decline inmortality rate. Literature survey reveals that early detection of the cancer plays a key role in the reduction of death rates. Generally, earlier the breast cancer detected, better the survival rate, because the treatment is more effective before the disease is spread.

Because the cause of breast cancer has not been understood yet, so, primary prevention is not possible. Currently available treatment methods are very effective in it at early phase. Therefore, at early stage removal of cancer is a promising way to achieve the significant increase in survival rates [14, 67].

As mentioned in Section 1.1 of this dissertation, according to American Cancer Society (ACS), between 1973 to 1999, incident rates of breast cancer have increased by nearly 40%, however, between 1989 and 1995 the mortality rates were declined by 1.4% per year and 3.2% afterwards, due to early detection of the disease [3]. Similarly in Europe, during 1985 to 1989 a reduction of mortality rate was 2.1%, by the availability of advanced imaging modalities [7].

A study regarding breast cancer death rate was performed among the Swedish women, during a time span in which the government provided breast screening facilities widely available to the women. The death rate was dropped 63% for the women who had adopted periodic breast screening procedures, as compared to 10 years before when such screening facilities were not available [68].

A number of advanced imaging modalities are being used for breast screening to diagnose and detect the cancer at early stage. Some of them are discussed below:

**Mammography**

It has been a challenge to devise an imaging device which may detect breast abnormalities and determine whether they are malignant or benign accurately at early stage. Mammography is the most common breast imaging technique which is commonly used forthis purpose. It is considered as the essential tool to detect the said cancer at early stage andwith smaller size lesions, which in turn reduce the death rate due to this cancer [69]. It has been proved that breast cancer mortality rate is reduced by 30% (specially for the women with the age between 40-74 years), by the implementation of mammography [70-71].

Wilhelm Konrad Roentgen discovered X-rays in 1895 and for which he was granted first Nobel Prize for Physics in 1901. These rays have been in medical use since 1896. Mammography is also a use of X-rays for the breast screening. The first mammography systems which utilized X-rays for the imaging of breast tissues were developed in 1970s [72]. A schematic diagram of a mammographic system is presented in Fig. 2.6.

Now a days, two different X-rays mammography imaging techniques are available:

a) Screen film mammography (conventional mammography)

b) Digital mammography

**Screen Film Mammography**

Like other radiographic examinations, breast is exposed to low energy X-rays photons beam from one side, and the variation in beam intensity due to the absorption of beast tissues is recorded from the other side. Generally in conventional mammography the breastis compressed between a polycarbonate compression plate on the cranial side, and photographic film holder on the caudal side (Fig. 2.7).

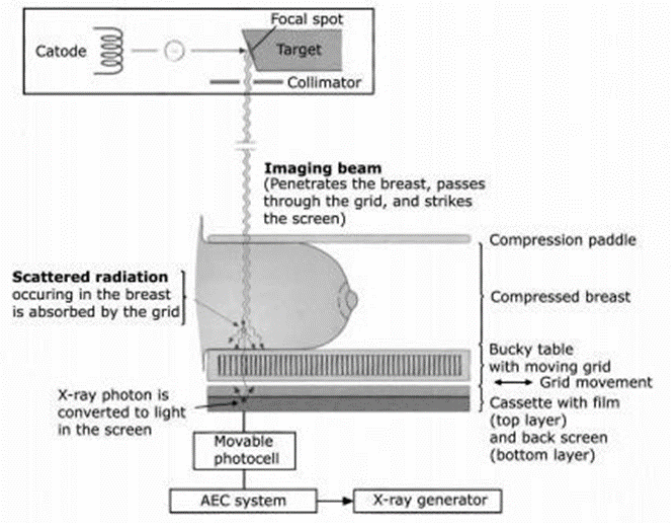


**Fig. 2.5:Schematic Diagram of Mammographic System [73].**

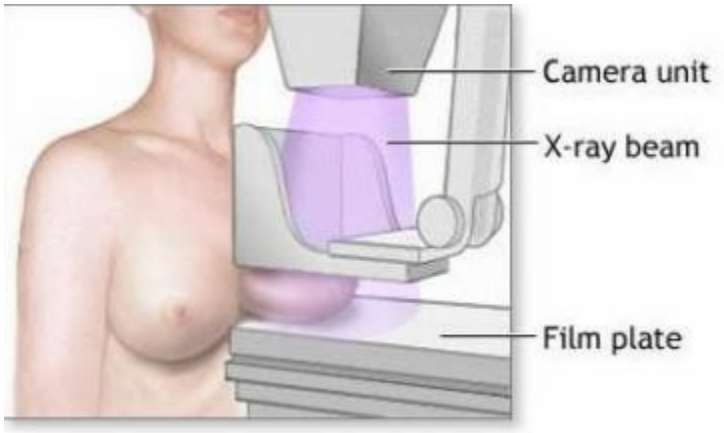



**Fig. 2.6:Mammography Procedures[74].**

Because different tissues of the breast absorb different amount of X-rays radiations, so, it makes possible to explore the details of the tissues to be examined. Image of the breast varies between individuals due to differences in relative amount of fat, connective and epithelial tissues, and because of the differences in radiodensities of these tissues. In mammographic images, fat appears dark because radiographically it is lucent, whereas, connective and epithelial tissues are radiographically dense, and appear light.

**Digital Mammography**

A film base system has a non-linear transfer function, because the photographic film sensitivity depends upon X-rays dose. This results in limited dynamic range making it difficult to detect tissue having high attenuation for X-rays, from the tissue with low attenuation in the same exposure. To prevent under or over exposure of the film, X-rays dose must be very carefully adjusted.

To overcome these problems film cassette is replaced by digital receptor in mammographic systems. First digital systems were based on the storage of phosphor, so called digital luminescence or computer radiography (CR), but, these systems have poor signal-to noise ratio at normal dose of X-rays. The main advantage of CR is the high contrast of the image. Now-a days silicium or selenium have been introduced as detectors in novel direct digital mammography systems. High quanta efficiency is one of the advantages of the digital systems, which provide higher contrast, detailed resolution at reduced dose. There isa linear relationship between dose and detector signal [21].

The main advantage of digital imaging system is the possibility of splitting the acquisition, processing, and display processes into separate stages, and offer the possibility of independently improving each stage. As the images are displayed immediately on the monitor which reduces the examination time significantly, allows for real-time stereotactic biopsy, and allows remote and prompt consultation with other physicians via internet, and enables to avail the benefits of digital processing such as image enhancement, image magnification, image archive and retrieval [72].

Moreover, digital imaging system permits the utilization of advanced computer applications such as computer aided detection/diagnostic (CAD) algorithms to assist radiologists in the differentiation of abnormalities as benign and malignant in mammograms (the same benefit of digital mammograms has been utilized to complete thisresearch work).

One perceived disadvantage of digital mammograms is its lower spatial resolution as compared to film. In practice, the effect of this lower spatial resolution is not observed relative to film, because, the better resolution of the film is not realized due to its lower contrast. It has been demonstrated at the American College of Radiology (ACR) phantom for which no film system is presently in use can image all of the objects in the phantom in one exposure, due to its higher contrast and dynamic range [57, 72]. It is also less reliable for the dense breast young woman and for the women who have undergone a surgical procedure because glandular and scar tissues are as radio-opaque as abnormalities [75].

Generally, mammographic imaging of each breast is performed in two poses:

a) An overhead view, which is called craniocaudal view, an example of such animaging procedure and resulting image is presented in Fig. 2.7.

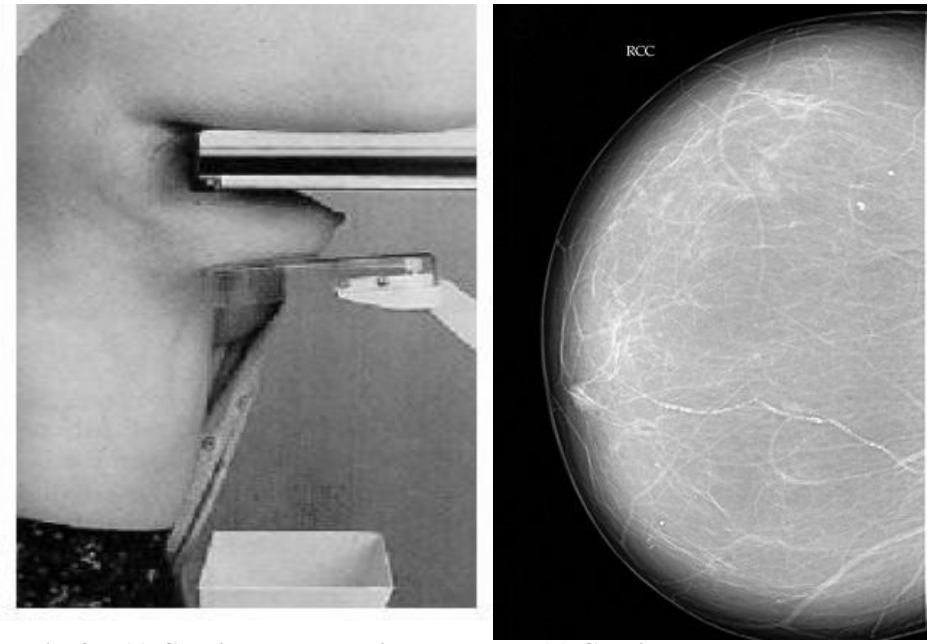



**Fig. 2.7:(a) Craniocaudel Imaging Procedure(b) Craniocaudel Mammogram**

b) Side view, which is called medioletral oblique, screening procedure and outputimage, is shown in Fig. 2.8.

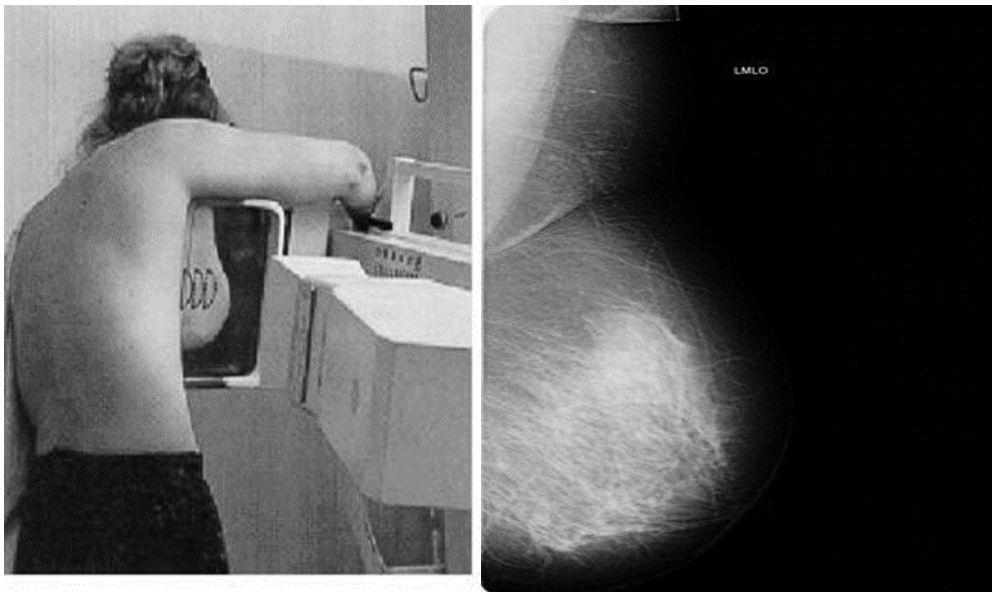



**Fig. 2.8:(a)Medioletral Oblique Imaging Procedure (b) Mammogram**

The different radiological appearances developed due to variations in the relative amountsof these tissues are termed as parenchymal patterns of the breast, which can mask the underlying cancerous lesions, making the interpretation of images difficult [76], so, the standard practice of taking two views of each breast has been has shown to be more sensitive and effective in detecting signs of cancer.

**Breast Anatomy in Mammograms**

Anatomical parts of a female breast which have been discussed in Section 2.1 are labeledin the Fig. 2.6.

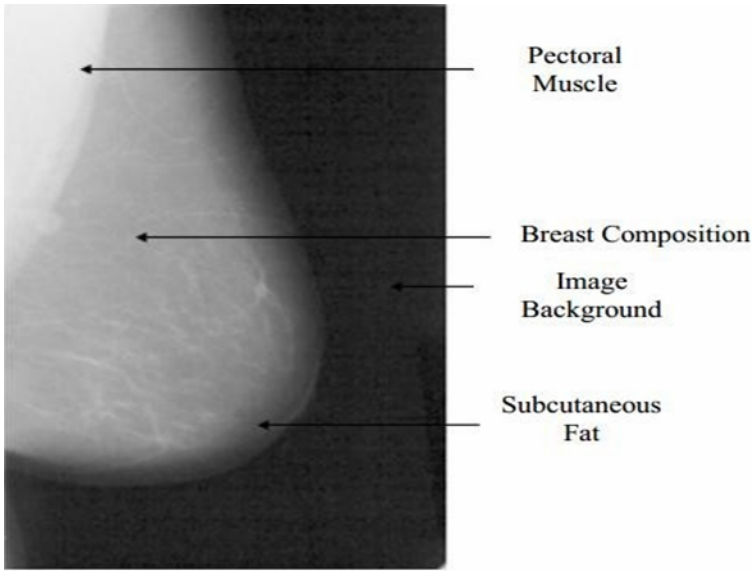



**Fig. 2.9:A Labeled Mammogram**

This figure shows a mediolateral mammogram in which the basic anatomical components are labeled. The denser tissue such as muscles, fibroglandular tissue, malignant and benign masses, and vascular tissue, having higher attenuation property for X-rays appear brighter, whereas, areas containing fat or skin appear darker due to lower attenuation property.

**Findings of a Mammogram**

In addition to the basic anatomy of the female breast, a number of other objects such as masses, microcalcifications, and architectural distortions, also appear in a mammographic image.

Mass is a space occupying lesion, it is referred to as a density if it is seen in only oneprojection. If it is viewed in both projections then it is called mass [77]. Several types of masses are found in mammograms, which are categorized on the basis of their shapes, densities, and margins. The shapes may be round, oval, lobular, and irregular. The margins include; microbulated, circumscribed, obscured, indistinct, and spiculated. The densities may be high density, low density, equal density, and fat containing. All these factors help radiologists to describe the masses found in mammograms and to classify them as benignor potentially malignant.

Masses of different shapes, which are usually appeared in mammograms, are shown in Figs. 2.10-2.13. Round masses appear in circular shapes and oval masses are displayed in elliptical shapes. Lobular masses are displayed as contours and undulations. The masses having undefined shapes are called irregular.

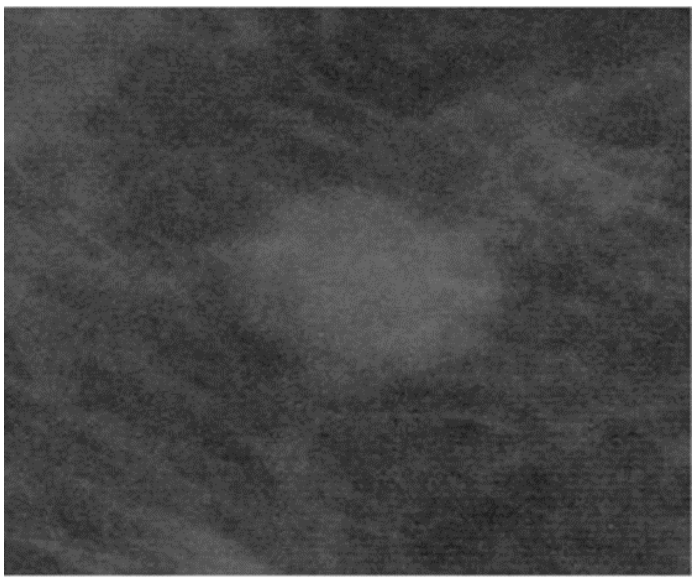


**Fig. 2.10:Mammogram with Round Mass**

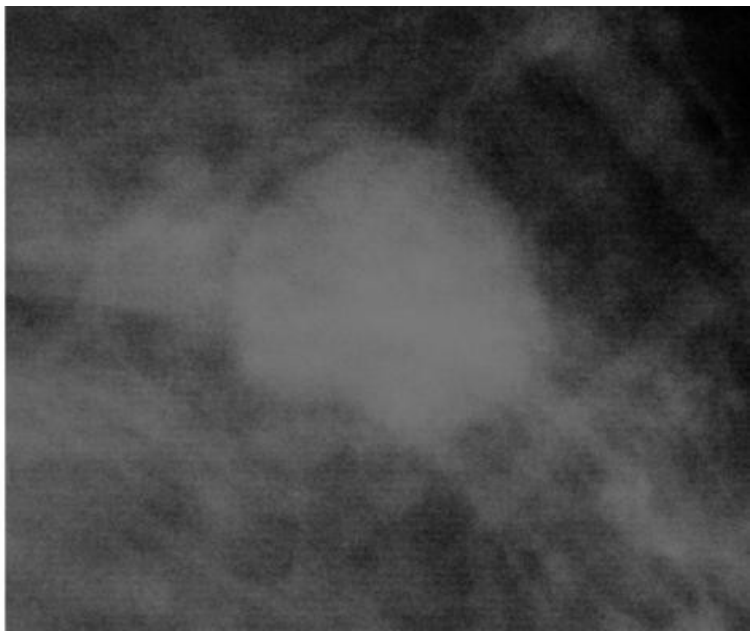


**Fig. 2.11:Mammogram with Oval Mass**

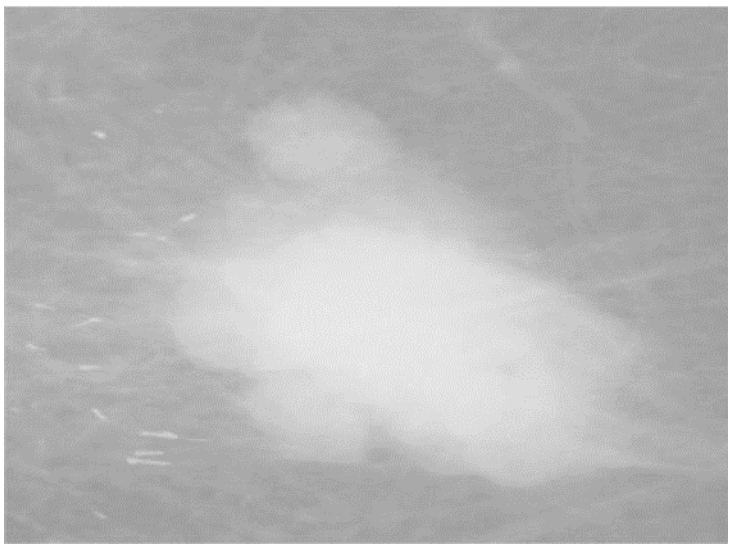


**Fig. 2.12:Mammogram with Lobular Mass**

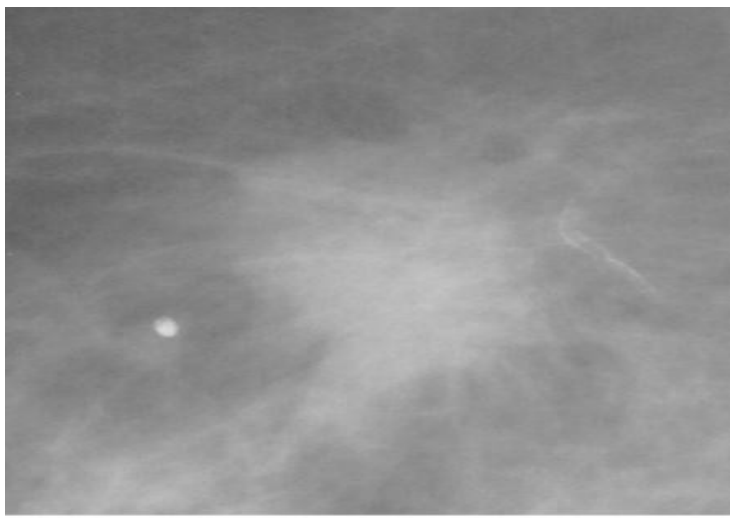


**Fig. 2.13:Mammogram with Irregular Mass**

Masses with varying types of boundaries characteristics are also observed in mammograms. Circumscribed masses are displayed with distinct well defined boundaries. Microlobulated have margins which are undulated in small cycles. The masses which are hidden by superimposed or adjacent normal tissue are called obscured masses. Indistinct masses have poorly defined boundaries which taper into the background. Spiculated masses appear spoke-like lines radiating out from the mass. Figs. 2.14-2.18 illustrate the varying types of boundaries masses.

**Fig. 2.14:Mammogram with Circumscribed Mass**

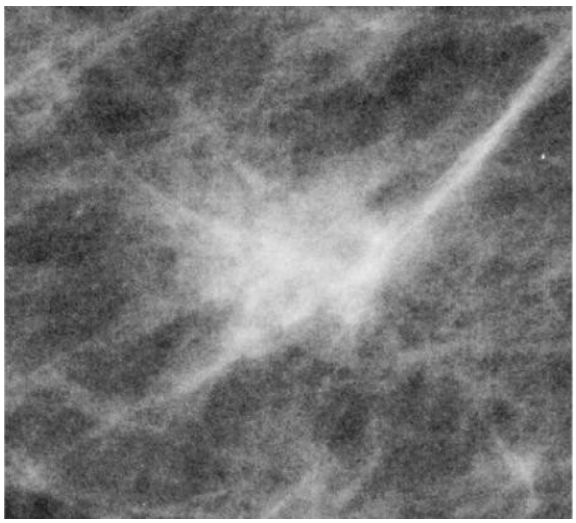


**Fig. 2.15:Mammogram with Microlobulated Mass**

**Fig. 2.16:Mammogram with Obscured boundaries Mass**

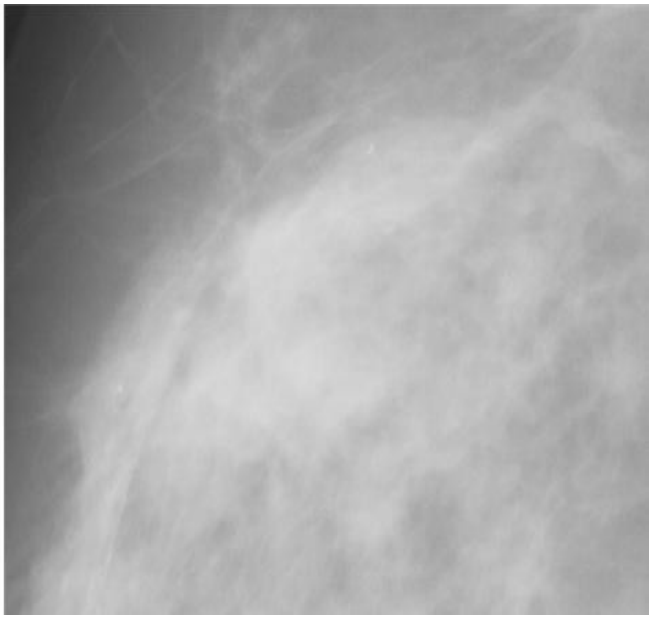


**Fig. 2.17:Mammograms with Indistinct boundaries Mass**

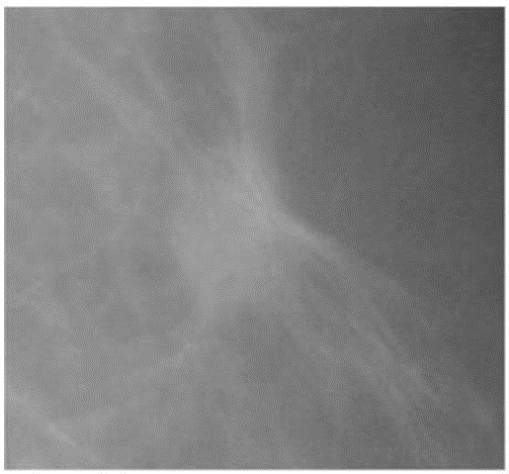


**Fig. 2.18:Mammogram with Spiculated Mass**

Varying density masses are also displayed in mammograms. Figs. 2.19-2.21 illustrate someexamples of varying density masses. The masses with high density stand out sharply from the background tissue, whereas, masses with low density are faint and such type of masses may be partially obscured. The masses having equal densities appear similar to other components in the mammogram. The masses which contain fat also appear faint due to lowdensity fat deposits in the mass.

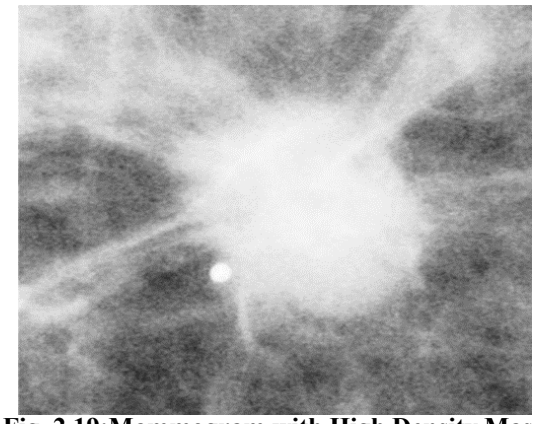


**Fig. 2.19:Mammogram with High Density Mass**

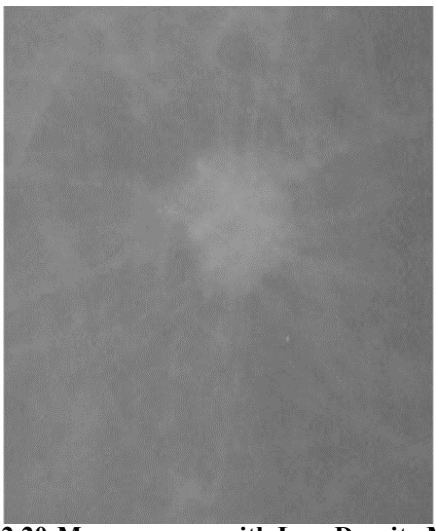


**Fig. 2.20:Mammogram with Low Density Mass**

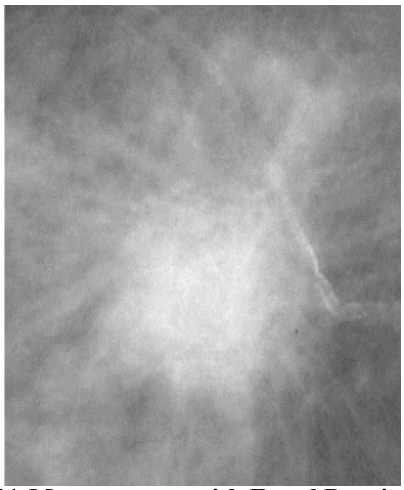
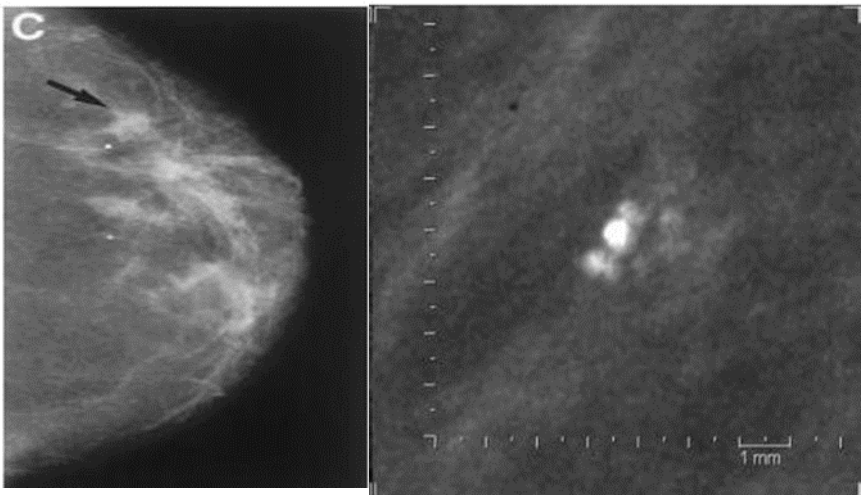


**Fig. 2.21:Mammogram with Equal Density Mass**

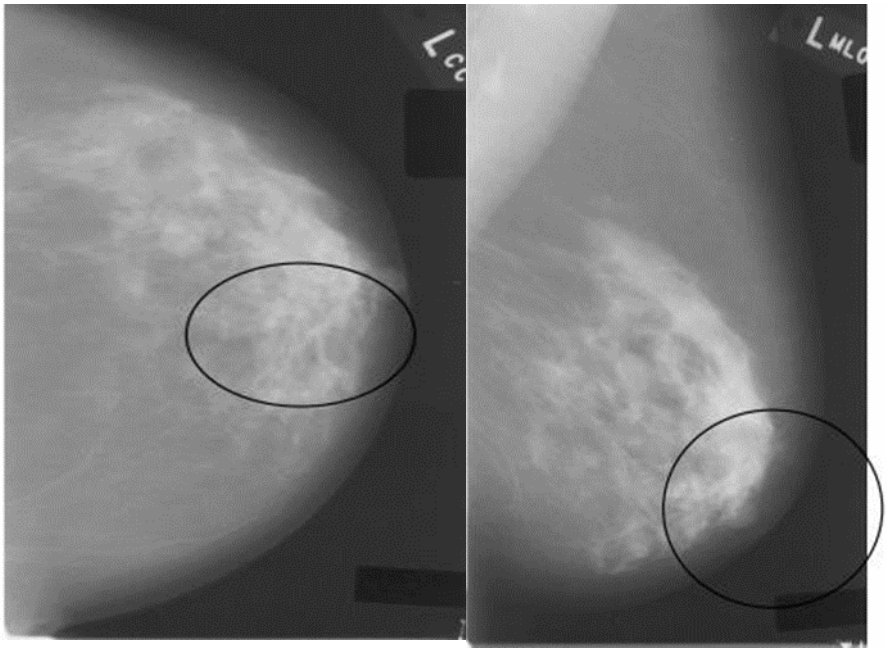
Other the most important objects which are displayed in mammograms are microcalcifications. A detailed discussion of microcalcifications has been already presented in Section 2.2. These are the small deposits of calcium in breast tissue, which areconsidered as highly indicative of breast cancer. The MCs appear as bright dots or spots in the mammograms. In Fig. 2.22 (a, b) illustrate the appearance of MCs in mammograms.




**(a)** **(b)**

**Fig. 2.22:Mammograms with Microcalcificaitons**

Another type of abnormality which may be observed in mammograms is the architectural distortion of the breast. Here normal structure of the breast observably distorted from a focal point. These distortions are seen in images even though no mass is observed there in that mammogram. Fig 2.23 (a, b) illustrates such type of abnormalities.



**(a)** **(b)**

**Fig. 2.23:Mammograms with; (a) side and (b) overhead view of Architecturalabnormalities.**

Some other phenomena are also observed in mammograms, which make it difficult to interpret for a radiologist, for example mammograms of dense breasts. They display fibroglandular tissue, which makes it difficult to detect cancerous signs. An example of such a case is presented in Fig. 2.24.

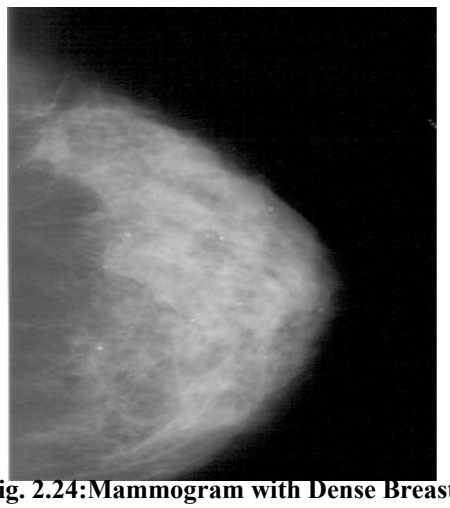


**Fig. 2.24:Mammogram with Dense Breast**

A number of other imaging modalities are also being used for the diagnosing or detectionof breast problems. Some of them work in isolate environment and some other are employed as adjunct to other modality. A brief description of some well known imaging modalities is presented here.

**Magnetic Resonance Imaging (MRI)**

Powerful magnetic fields and radio waves are used to create images in MRI. A large cylindrical magnetic is the main component of this system. In imaging process a radio signal is turned on and off, and subsequently, a part of these waves are absorbed by the molecules of the body, and a part of these waves is echoed by the body. These echoed waves are continuously recorded by a scanner of the system and a computer

converts these signals into image.

One of the benefits of MRI is that with this system image of the breast can easily be acquired in any orientation, but, this provision is not available in other modality. In MRI a contrast enhancing material is injected into the vein of arm, because, without contrast enhancement this tool for breast imaging has not been proved useful in the detection or diagnosis of said cancer. Previous studies verified that contrast enhanced MRI imaging has been a sensitive modality for the detection of tumor recurrence, with sensitivity of nearly 100% [78]. Contrast enhanced MRI images may be very helpful in determining the extent of breast cancer as a management aid, detecting suspected foci of multifocal or multi- centric tumors and perhaps helpful to differentiate malignant and benign lesions.

There are some abnormalities which can only be detected by MRI but may not be with mammography or ultrasound especially in the dense tissues, making an MRI guided system necessary [79]. It has shown a valuable screening method for the women with high risk of breast cancer, and therefore it has been included in the American Cancer Society (ACS) recommendations for screening [80].

No doubt, MRI may be a promising supplementary tool to mammography in the diagnosing of breast cancer, but, there are some hurdles to use MRI in this regard. While mammography can image microcalcifications but MRI cannot image these tiny deposits. Also, MRI cannot always differentiate cancerous and non-cancerous breast abnormalities, which leads to the un-necessary breast biopsies. MRI screening is costly and time consuming as compared to mammography [72].

**Computer Tomosynthesis (CT)**

The major problem regarding the detection of small cancers is the overlapping of tissues ofthe breast. It is termed as structure noise. The major advantage of tomography is the reduction in this structure noise [57]. Conventional computed tomography would not be suitable in this regard because head and body of the patient would interfere with the required movement of the detector and X-rays source. However, tomography can be achieved with the breast held in the standard mammographic system and with the static detector. This is the basic principle of digital tomosynthesis. In this system images are taken at low dose with a limited range of angels X-rays source, and digitally shifts them so that only structures in the same plane align in the combined image. Because of misalignment structures in the other planes are blurred, so, in this way visibility of the structures in the desired plane is enhanced.

Recently, this modality has been mainly used to evaluate mediastinal nodes in oncology. Because this technique uses the size as main criterion to assess the status of modal, so, the lesions smaller than 1cm are usually described as non-cancerous, due to their small size, and also metastatic lymph nodes are often not identified by CT [81]. Rate of breast cancer recurrence detection by using CT, in patients without a palpable axillary mass is extremely low, so, it is suggested this technique for breast screening is not justified for the clinically occult axillary disease in patients with arm symptoms following axillary surgery or radiation therapy for breast cancer.

**Ultrasound Tomography (UT)**

In this modality high frequency sound reflected waves are used to produce image of hiddenstructures that have characteristically different densities. This technique of imaging is not used as a substitute of X-rays mammography, but it is best to differentiate the cystic from the solid lesions and as a guide for aspiration and biopsy [57]. In fact UT is an essential complement to physical examination and mammography to evaluate breast masses.

One unusual characteristic of the breast is that relative to parenchyma, fat present in the breast is hypoechoic, which is the major problem regarding to ultrasound tomography because, with rare exceptions all breast cancers are also hypo-echoic [57]. Some significant breast cancers are difficult if not impossible, to see using this technique becausethey are iso-echoic with fat or breast tissue.

**Positron Emission Tomography (PET)**

Positron Emission To mography with fluoro-2-deoxy-D-glucose (FDG) is based on the changed metabolic properties of cancerous tissue. The intake demand of cancerous tissue for this chemical is increased as compared to normal tissue. PET imaging for whole-body analysis to evaluate the majority of tumors has raised interest in its use for the diagnosis of primary breast cancer [82].

FDG PET seems to have a reasonable sensitivity and specificity in the detection of recurrent and metastatic breast cancer, especially for the patients presenting with elevated tumor markers [82].

Although whole-body FDG PET has a certain diagnostic accuracy for the detection of malignant breast lesions, but its sensitivity is lower than other standard diagnostic modalities. Moreover, PET application is limited due to requirement of radioactive material fluoro-2-deoxy-D-glucose (FDG), that is difficult to produce because this materialis made using on site cyclotron due to its short lifetime [82].

**Thermal Imaging**

This method of breast imaging is based on the assumption that development of cancerous breast lesion causes increase in blood flow, which can be detected by highly sensitive thermal sensor. This sensor operates in mid to long wavelength range of infrared region of spectrum of electromagnetic waves. Early attempts for this technique were poorly receivedand have given the bad reputation, because the early models of thermal sensor werecharacterized by noise equivalent temperatures (NEΔT) of the order of one degree centigrade. Technically the system based on photovoltaic InSb focal plane arrays and GaAs quantum well infrared photo-detectors (QWIP) have improved NEΔT to the order of .025$^{o}$ centigrade.

The main advantage of this technology is the low cost of system and lack of any ionizing radiations. The most significant disadvantage of this modality is that, due to the very high emissivity of tissue from water absorption, only thermal signatures on the surface of the skin can be detected. So, pathological lesions on the surface of the skin easily be detected but the deep lesions would be substantially screened by the insulating properties of the breast tissues and by the diffusion of the heat in reaching the surface. Hence, this technology may play a role in breast imaging, but could never be considered as a primary tool.

**Electrical Impedance Tomography (EIT)**

Electrical Impedance Tomography (EIT) is based on the distribution measurement of electrical impedance in a cross-section of the breast. This is possible because different tissues have different electrical resistivity. A series of small current pulses are applied to the breast, and a set of potential difference measures are made from the non-current carrying electrodes. Because the applied current flow though the least impedance paths, depends upon the subject‟s conductivity distribution [83].

It represents a potential, non-invasive technique to image the breast cancer. Ex-vivo studieshave shown characteristic frequency changes in the electrical conductivity and permittivity of cancerous versus non-cancerous tissues. By applying a current around the breast periphery and by measuring the resulting voltage, one can use computational model-based approaches to develop tomographic images.

In future this technique may be a powerful new modality in the arsenal of diagnostic mammographic techniques.

**Computer Aided Diagnostic/Detection (CAD) System**

Currently, it is assumed that mammography is the most reliable modality for the detection of breast abnormalities. Conventionally, the radiologists classify these abnormalities on the basis of some visual parameters (qualitative parameters) like shape, size, contrast, border conditions *etc*. Subjective nature of all these described parameters results the misclassification/misinterpretation of abnormalities present in mammograms. Analysis consistency also suffers due to nature of data evaluation which depends upon the training, experience, and fatigue of the radiologists. All these factors lead to inter/intra observer variability among the radiologists for the interpretation of same mammogram [84-86].

In such an environment various researchers have employed artificial intelligence by using computer to help the radiologists in the decision-making process regarding to the interpretation of abnormalities, as malignant or benign. Since last two decades the investigators have focused on the development of computer aided diagnostic/detection (CAD) tools for the beast cancer. The average sensitivity of a radiologist may be increased by 10% with the implementation of CAD system [87]. Literature survey demonstrates that 12-19.5% missed breast cancers would be detected by the use of such CAD system [88]. The accuracy of advanced CAD system is 100% for the detection of breast tumors in early stage [89], and such detection is a fundamental sign for accurate diagnosis and disease management [34]. CAD system plays a role of a virtual „second reader‟ for the radiologiststo make the final decision [19].

Regarding the CAD system for the radiological image interpretation Chan *et al.,* in 1995 expressed as:
―*Although a computer program may never be able to achieve the level of knowledge and cognitive capability of radiologist, a trained computer program canperform certain tasks reproducibly and consistently without inter-observer and intra-observer variations that are commonly observed among human observers. The ability of a CAD scheme can therefore be complementary to that of a radiologist*‖ [90].

Studies have proved that CAD techniques have the potential to improve the diagnostic accuracy and reduce the use of adjunctive procedures morbidity, as well as cost of treatment. This technique can facilitate the enhancement, detection, characterization, and quantification of diagnostic features, such as shape of MCs and masses, development of tumors into surrounding tissues, and distortions due to developing densities [91].

For large-scale screening applications, CAD methods are important for:
Prototyping robust systems working independently of patients.

Improving sensitivity in cancer detection.
Reducing observer variation in image interpretation.
Facilitating remote reading by experts (*e.g*., tele-mammography).

CAD techniques are being used for the detection of suspicious regions in the mammograms, segmentation of mammograms, to enhance the contrast of mammographic images, for the classification of abnormalities present in mammograms *etc*. In this work we have also employed CAD technique for the classification of breast cancer by analyzing suspicious regions in digital mammograms on the basis of quantative parameters.

**Previous Work**

As mentioned in previous section that CAD system is being employed since last three decades, by a number of researchers for the detection, diagnosing, and classification of breast abnormalities in mammograms. In such a system quantitative descriptors (textural features or parameters) are extracted from the specific regions in the image and these descriptors are deployed to an artificial intelligence algorithm which helps to separate the normal and abnormal tissue by pattern recognition procedure.

Chan *et al.,* employed such a CAD system to investigate whether the existing microcalcifications are associated with benign or malignant pathology, in which 86 mammograms were used to extract statistical features. Up to 10 pixel distancetotal 13 textural features were calculated on the basis of SGLDM in axial and diagonal directions. Stepwise feature selection algorithm was used to select the most appropriate set of features. Back propagation neural network was applied for training, and for testing leave-one-out algorithm was implemented to recognize the normal and pathological microcalcifications. The system efficiency was estimated by ROC, on the basis of six textural features an accuracy of 88% was claimed [92]. A very simple method based on CAD has been used by Vasantha *et al.,* to differentiate benign and malignant masses in mammograms, taken from MIAS database. The investigators extracted 26 textural features (6 first-order Histogram parameters and 20 second-order GLCM parameters), from the specified ROIs. Prior to features extraction histogram equalization was employed to enhance the images contrast. The most relevant features selection was performed by a combined scheme of GA algorithm and stepwise reduction method. For the classification of masses as normal and pathological, decision tree classifier was employed. With this approach an accuracy of 95% was received by the authors [93]. In a research work conducted by Yu and Guan, Wavelet coefficients and GLCM features were extracted from the mammograms selected from Nijmegen dataset. The most discriminative features were selected by applying General Regression Neural Network (GRNN), with this algorithm a set of nine features was selected. The selected features were subjected to the automatic detection of clustered microcalcifications. The system accuracy was evaluated by Free Response Operating Characteristics (FROC). The proposed CAD system achieved an accuracy of 75% at the cost of 0.5 false positive per image [61]. The same researchers in another work used 31 same types of features for the detection clustered microcalcifications. The discrimination ability of these features was analyzed with GRNN by sequential forward and sequential backward selection technique. Multilayer feed-forward neural networks were implemented, and 90% system accuracy was achieved [94]. Schaefer and his co-authors proposed a CAD system on the basis of pattern recognition for thermographic breast images, where they used first order and GLCM textural features. The most relevant feature set was selected by Mutual Information (MI) algorithm, and fuzzy rule based classifier was implemented to classify benign and malignant tumors. They achieved a classification accuracy at rate of 79.53 percent [95].

Breast tumors in UT images were categorized by Cben and Liu, with the implementationof CAD method, on the basis of GLCM features. Input data was classified by Fuzzy C- means and k-mediod clustering algorithms, and experimental accuracy was 72.64% [96]. Shi *et al.,* also worked on UT images for the detection and classification of breast masses. They used fractal textural features in their work and employed SVM classifier for differentiation of masses as normal and pathological and obtained an accuracy of 96.4% bythis work [97].

On the basis of statistical textural features (first order + Haralick"s parameters) and waveletcoefficients microcalcifications were classified by Kulkami*et al.,* and the authors used Probabilistic Neural Network (PNN) for classification the of microcalcifications (as normaland pathological), whereas, ROC was employed for the assessment of system accuracy [67].

Rejani and his companion, proposed a pattern recognition procedure for the classification of breast tumors. Discrete Wavelet Transform (DWT) was employed to extract various texture features from the segmented images. Tissue classification was performed by SVM and received 88.75% accurate results [98]. Shape and texture descriptors (GLCM features) were implemented by Matrins *et al.,* to classify segmented regions in digitized mammograms in two groups; mass and non-mass. SVM classifier was employed for the separation of mentioned classes with CAD system. The researchers reported 85% accuracyin their results [99].

An automatic diagnosis system based on pattern recognition for breast cancer detection was proposed by Karabatak & Ince, where they employed association rules (AR) and ANNfor the reduction of input features and pattern classification. A total number of 9 descriptors were calculated from the dataset and by using AR method only 3 were selected as the most significant features, which were deployed to classifiers. The authors also compared the performance of ANN method with the performance of combination ofAR+ANN classifiers. They obtained 95.6% accuracy when ANN+AR scheme was employed [100].

Periera *et al.,* used CAD system to differentiate the abnormalities present in ROIs, in following four modes; (1) normal and abnormal, (2) microcalcification and masses, (3) benign and malignant microcalcifications, and (4) benign and malignant masses. For this purpose they extracted 13 (GLCM) statistical features in radial and axial directions, and 6 wavelet spectral features. The most relevant features were selected by Jeffries-Matusita distance algorithm and the classification was performed by k-NN classifier. The system efficiency was verified by ROC, which has been reported as for normal regions and lesions 95.7% and for masses and microcalcifications as 86%, but its performance is very poor for the differentiation of benign and malignant lesions which was 17% [101]. Differentiation of malignant tumors in mammograms was performedby Mu *et al.,* with the implementationof CAD scheme, on the basis of morphological, edge sharpness, and statistical features. Total 22 descriptors, including 14 GLCM features, 5 shape features, and 3 edge sharpness features, were extracted from 111 ROIs. The best descriptors were selected by GA based on different criteria such as class separability, normalized distance, and alignment of kernel with target function, whereas, classification of tumors was performed by *Fisher's LDA* (FLDA), SVM, and strict two-surface proximal (S2SP). The output efficiency is reported 95% [102]. In the same way Papadopoulos and his co-authors used morphologicaland textural features to detect microcalcifications in two sets of mammograms taken from Nijmegen and MIAS database. Prior to feature extraction images contrast was enhanced to locate potential microcalcification objects, and then morphological descriptors were employed to extract ROIs. A total 54 features were calculated from each ROI and for the selection of the most appropriate features ROC curve was plotted for each feature, so, a set of 22 features was selected. For classification three algorithms rule-based expert systems, ANNs, and SVMs were utilized. For Nijmegen dataset the classification by all three methods was 52%, 77%, and 81% respectively. Similarly, for MIAS dataset the classification was 61%, 78%, and 74% respectively. In the last the researchers also used *PCA* method to further enhance their results [103].

Zadeh *et al*., used 103 regions having microcalcification clusters in their work to extract shape, GLCM, wavelet, and multi-wavelet texture descriptors. Real-valued genetic algorithm (GA) and binary GA were applied to select the most appropriate features.Classification of malignant and benign MCs was performed by k- nearest neighbor (k-NN) classifier, whereas, ROC technique was implemented to estimate the system efficiency. They received 84% to 89% accuracy by real-valued GA and 83% to 88% by binary GA [104]. Costa *et al.,* applied CAD system to discriminate masses present in, Digital Database for Screening Mammography (DDSM) images. They used Principal Component Analysis *(PCA),* Gabor Wavelet, and (an efficient coding technique) Independent Component Analysis (ICA) methods for the extraction of statistical features from the mammograms, *LDA* for the discrimination of masses as malignant and benign and ROC was utilized to estimate the system efficiency. According to authors the results received by Gabor Wavelet, *PCA*, and ICA were 85.28%, 87.28%, and 90.07% respectively, which shows ICA method is the for the discrimination of masses [105].

To classify the masses in mammograms Zheng extracted 36 morphological and intensity distribution based features on 400 digitized mammograms by the implementation of CAD system. Prior to the features extraction he applied difference-of Gaussian (DOG) filter and threshold algorithms to highlight the suspicious regions in the mammograms. For the classification of obtained regions, as malignant or benign the author used ANN and k-NN classifiers separately and received an accuracy of 89% and 85% respectively. To have better performance the author used an algorithm in which both classifiers were combined and obtained an accuracy of 91% [106].

A hybrid classifier, which is the combination of supervised and unsupervised classifiers, was applied for the classification of malignant and benign masses on 348 ROIs in mammograms by Hadjiiski *et al.,* The statistical textural features were extracted from GLRM and GLDM in all directions up to d=20. The performance of designed classifierwas compared with *LDA* and back-propagation neural network (BPN) and it was 81% for hybrid classifier, 78% for LDA, and 80% for BPN [107]. To classify the microcalcification masses as benign or malignant a pattern recognition method was proposed by Mousa and his co-workers, on the basis of Wavelet texture features. In their work Fuzzy-Neuro classifier was implemented to have better results [108]. Similarly ColorWavelet textural features extracted from ROIs, were deployed by Niwas *et al.,* to SVM classifier and they obtained 98.3 percent accurate results [109].

It has been demonstrated through literature survey that the textures of breast tissue surrounding MCs is also very helpful for the diagnosis of cancer. Working in this way an active contour model or snake model was used by Liu *et al.,* to develop a band region around a suspicious spot and GLCM features for d=1 in all four

directions were calculated from this region. These features were deployed to linear discriminant analysis *(LDA)* and support vector machine (SVM) classify the masses as benign or malignant. The system efficiency is assessed by ROC, which is reported approximately 70% [110]. Similarly Karahaliou *et al.,* also diagnosed 100 MC clustered form DDSM database by analyzing thetexture of MCs surrounding tissue. The feature selection algorithms were based on exhaustive and LDA methods. The authors claimed an accuracy of 89% [111]. In the same way Zyout et al. worked on MC clusters from mini-MIAS and local dataset. They characterized the malignancy of MCs‟ clusters using GLCM features, extracted from surrounding tissue integrated into an embedded feature selection method based on a heuristic particle swarm optimization (PSO) k-NN approach. The reported accuracy is 100% for mini-MIAS dataset and 94% for the other dataset [112]. Similar to previous authors Kim and Park presented a comparative study of different texture analysis methods, using surrounding region dependence scheme to highlight the suspicious spots in mammograms. They extracted statistical textural features by GLCM, GLRM, and GLDM methods and employed three layer back-propagation neural network for classification, while, ROC was used to estimate the performance of the system. The authors received 93, 88, and 74 percent accuracy respectively with all above mentioned textural features [113].

Khuzi *et al.,* have applied CAD system for the identification of normal and abnormal masses. They used 40 images (20 with benign and 20 with malignant masses) fromMammographic Image Analysis Society (MIAS) database. Four GLCM texture descriptorsin all directions $0^o$, $45^o$, $90^o$, $135^o$ for d=1 were utilized for the discrimination. For classification threshold, k-NN, and Otsu‟s methods were employed, and ROC technique was implemented to verify the performance, which is reported 82%, 70%, and 84% respectively [19].

Masmoudi *et al.,* have categorized the breast densities by using machine vision approach according to the criterion defined by BIRADS system. 400 images (200 for training and 200 for testing) were used in this work. Local Binary Pattern Variance (LBPV) features were deployed to ANN classifier for training and testing the system, while performance of the developed system was checked by Receiver Operating Characteristics (ROC). They achieved an accuracy of 79% [114]. Similarly, Jele and his co-workers suggested a methodfor the automatic malignancy grading for needle biopsy breast tissue. With the use of SVM classifier they received 94.24% accuracy with their proposed method [115].

Fatima and Amine implemented CAD approach to classify breast cancer in Wisconsin Breast Cancer Diagnosis database by using Adaptive Neuro-Fuzzy Inference System (ANFIS). The said classifier was used on tissue level and an accuracy of 98.25% was obtained [116].

Loukas *et al.,* have utilized CAD system for the grading of breast cancer (I, II, and III grade) on the basis of shapes and architectural patterns of breast tissues. In this study a set of 60 images of cancerous tissues were used and for analysis 30 statistical textural features were employed. Three pattern recognition methods, k-nearest neighbor (k-NN), probabilistic neural network (PNN), and support vector machine (SVM) were utilized for classification of input data. A discrimination efficiency of all three classifiers in training phase was nearly 97%, 95%, and 97% respectively, whereas, for testing the accuracy was 86%, 85%, and 90% respectively [2].

Recently, Zhang and his co-investigators applied CAD technique to classify the tissues at cell level by using biopsy sample of breast and retinal images. They applied shape, GLCM, and LBP features for this purpose. *PCA* was implemented by using MATLAB 7 for each type of dataset, then a subset of features was selected by applying a combine rule and system efficiency was estimated by ROC. In this way they obtained an average classification rate of 92.28 percent for all types of classes [117].

## III. Results, Discussions, And Conclusion

The aim of this study is to develop a CAD system which based on the descriptors having objective nature, rather than subjective nature. The use of texture analysis for the classification of abnormalities present in mammograms might provide a scientific approach to standardize and to obtain the required accuracy in the diagnostic procedure. The schematic diagram Fig. 1.2 displays a general structure of the proposed CAD system. The detail of experimental work regarding the classification of two types of digital images, with the application of proposed approach has been presented in Chapter No. 4 and 5. The classification results of the presented work have been discussed the following section.

To classify the masses and lesions by a radiologist the borders, shape and relative intensities of this region are important descriptive features. The BI-RADS (already discussed in Chapter 2) also provides a mechanism to characterize a given abnormality in amammogram which also based on shape assessment of the mass or lesion and modified boundaries of the mass. All these mentioned features are subjective in nature.

**Classification results of test data**

In Chapter No. 4, the proposed system is applied to classify (differentiate) the five types ofwheat kernels named, Aas, Bhakhar, Farid, Meraj, and Punjnad. The main objective of thispart of work is to find out the settings for which the proposed approach has shown the bestperformance. The adopted procedural steps are briefed below:

Region of interests (ROIs) with window size (8x8), (16x16), (32x32), and (64x64)are developed in the input data images.
Two normalization conditions, μ± and1%-99, are applied prior to featuresextraction.
Under these normalization conditions following five types of statistical textural parameters: (1) 9 First-order statistical parameters, (2) 11 Second-order statistical parameters on the basis of GLCM, (3) 5 Higher-order statistical parameters from GLRM, and (4) 5 Auto-regression parameters are computed from each ROI.
For classification, the most significant parameters for each size of ROI are selected by *F, POE+ACC,* and *MI* approaches individually, and a combined set of features is also selected by merging the parameters obtained by these three mentioned techniques for both normalization conditions.
In the last step, by using three approaches, *PCA*, *LDA*, and *NDA* classification (dataanalysis) is performed on the basis of these selected features for all size of ROIs and under both normalization approaches.

For the classification of test image data, in first attempt, the most relevant features are selected by Fisher Co-efficient (*F*) approach from all size of ROIs. Classification performance on the basis of these selected features is checked for each size of ROIs by the implementation of all above mentioned multivariate analysis approaches (*PCA, LDA,* and *NDA*). The classification results are summarized in Table 6.1.

**Table 6.1: System performance with the Fisher features selection approach**

| Normalization | ROI | Classification (%) | | |
|---|---|---|---|---|
| | | PCA | LDA | NDA |
| μ±3σ | 8x8 | 51.53 | 57.87 | 68.00 |
| | 16x16 | 69.47 | 68.27 | 80.03 |
| | 32x32 | 79.88 | 82.35 | 89.32 |
| | 64x64 | 89.13 | 88.15 | 89.28 |
| 1-99% | 8x8 | 53.13 | 54.73 | 64.00 |
| | 16x16 | 68.67 | 67.27 | 69.27 |
| | 32x32 | 83.03 | 83.80 | 89.19 |
| | 64x64 | 82.47 | 92.56 | 96.13 |

It can be observed from the results presented in Table 6.1, that system has shown the best performance of 96.13% for ROI (64x64) by *NDA* analysis approach, when input data is normalized with 1-99% technique. The results acquired with this feature selection are not satisfactory, so to have better results another feature selection method, Probability Of Errorand Average Correlation Co-efficient (*POE+ACC*), is employed for all size of ROIs under both normalization conditions, as second attempt. To verify the classification results on the basis of these selected features the same three multivariate analysis approaches employed. The received system performance with this approach is presented in Table 6.2.

**Table 6.2: Classification with POE+ACC feature selection approach**

| Normalization | ROI | Classification (%) | | |
|---|---|---|---|---|
| | | PCA | LDA | NDA |
| μ±3σ | 8x8 | 33.63 | 38.27 | 25.73 |
| | 16x16 | 49.13 | 52.40 | 59.20 |
| | 32x32 | 77.72 | 78.12 | 84.57 |
| | 64x64 | 85.17 | 87.29 | 88.31 |
| 1-99% | 8x8 | 37.67 | 31.20 | 31.27 |
| | 16x16 | 53.80 | 61.60 | 73.27 |
| | 32x32 | 75.83 | 73.17 | 85.80 |
| | 64x64 | 80.47 | 83.07 | 89.87 |

It is evident from the results presented in Table 6.2 that with this features selection approach, our system has shown the best performance of 89.87% when *NDA* analysis approach is employed for ROI (64x64), which is lower than the previous scheme in which features are selected on the basis of Fisher Co-efficient presented in Table 6.1.

So, third attempt to improve the classification performance of the system is made by the implementation of Mutual Information (*MI*) feature selection approach. Similar to already adopted method, the most significant features are selected from all size of ROIs, and for classification, these selected parameters are subjected to already implemented three multivariate approaches. The performance of the system on the basis of these features, withall employed analysis approaches, for all size of ROIs and under both normalization techniques, is presented in Table 6.3.

**Table 6.3: Classification with MI feature selection approach**

| Normalization | ROI | Classification (%) | | |
|---|---|---|---|---|
| | | PCA | LDA | NDA |
| μ±3σ | 8x8 | 51.20 | 54.43 | 37.40 |
| | 16x16 | 69.17 | 68.67 | 78.20 |
| | 32x32 | 84.37 | 84.17 | 93.35 |
| | 64x64 | 91.40 | 90.17 | 97.85 |
| 1-99% | 8x8 | 33.13 | 54.73 | 36.67 |
| | 16x16 | 67.60 | 72.00 | 79.40 |
| | 32x32 | 82.07 | 83.40 | 90.00 |
| | 64x64 | 93.60 | 91.20 | 97.47 |

In this scheme we received the best performance with an accuracy of 97.85%, again, by *NDA* approach and for ROI (64x64), but under μ±3σ normalization conditions.

The classification results of all above implemented approaches are less than the expected value, so, finally a hybrid model of feature selection is employed. Here, a set of the most significant features is selected by merging the features selected by individual, *F, POE+ACC,* and *MI* approach, from all size of ROIs, under both normalization conditions. Again in this attempt, the system performance on the basis of this combined set of features is verified by three multivariate analysis approaches, *PCA, LDA,* and *NDA*, and the output results are summarized in Table 6.4.

With the implementation of hybrid feature selection model we received the excellent classification result with an accuracy of 99.67% by *NDA* approach from ROI (64x64) under 1-99% normalization condition.

**Table 6.8: Classification with combined set of feature selection approach**

| Normalization | ROI | Classification (%) | | |
|---|---|---|---|---|
| | | PCA | LDA | NDA |
| μ±3σ | 8x8 | 52.13 | 53.93 | 46.47 |
| | 16x16 | 69.67 | 75.47 | 86.20 |
| | 32x32 | 86.91 | 89.93 | 96.37 |
| | 64x64 | 89.11 | 91.13 | 97.98 |
| 1-99% | 8x8 | 49.80 | 65.53 | 31.87 |
| | 16x16 | 64.77 | 73.60 | 86.47 |
| | 32x32 | 83.63 | 84.37 | 96.13 |
| | 64x64 | 93.77 | 96.00 | 99.67 |

As per limit of the image size, it is impossible to manage a sufficient amount of sub-image data to achieve reliable statistical results. Hence, the regions of interests greater than window size (64x64) are not tried.

On the basis of parameters for which we have received the best results artificial neural network (*ANN*) is trained and tested. The input data of test images is classified with an accuracy of 99.90% and 93.11% in training and testing phase respectively (as presented in Section 4.4).

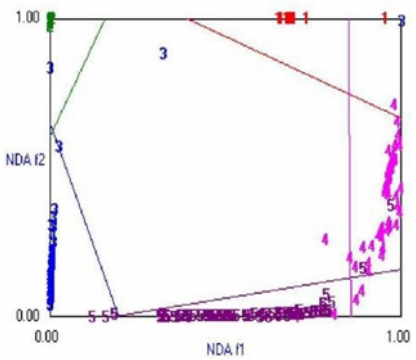



**Fig. 6.1: Clustering of data in NDA space for the best analysis**

**Classification results of experimental data (Mammograms)**

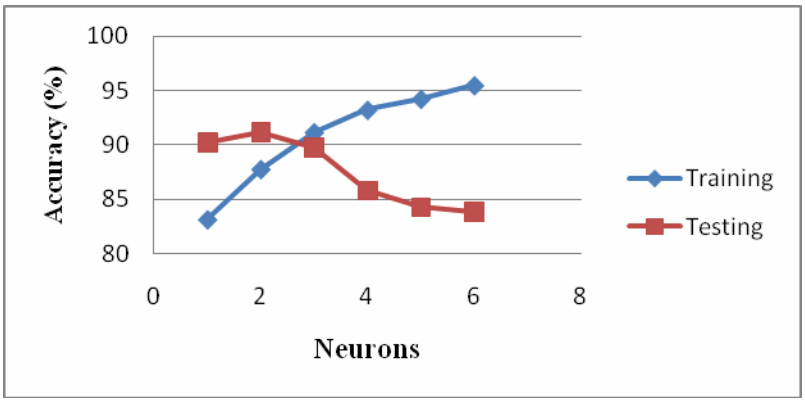
The validity and integrity of the proposed system is verified with the classification of test data. On the basis of results presented in Section 6.1, settings of the system, under both normalization conditions, for which

we obtained the best performance can be summarized as; (a) *NDA* approach for data analysis, (b) the data of large number of relevant texture features, and (c) large window size of ROI.

As we have already discussed in Chapter 4, that *NDA* approach based on *ANN* classifier and in B11 software a number of options are available to configure or tune up the said approach. To have better results the classifier is configured by applying different values learning rate „η" for different number of neurons in input hidden layer (for detail see Section5.3). To have more reliable results it is tried to use a smaller size of neural network as possible, which is measured by number of neurons in hidden layer [225], in this workup to 6 neurons has been tried. Similarly, only a small range for learning rate „η" (0.15- 0.30) is employed to have the best data clustering in *NDA* projection space, because, too high value of learning rate means, no learning at all [272].

To differentiate the abnormalities present in digital mammograms, as benign and malignant, the real time data has been analyzed by *NDA* approach under the above discussed architectural settings in Chapter 5.*Craniocaudel* (CC) and *Medioletral Oblique* (MLO) view images are analyzed separately.

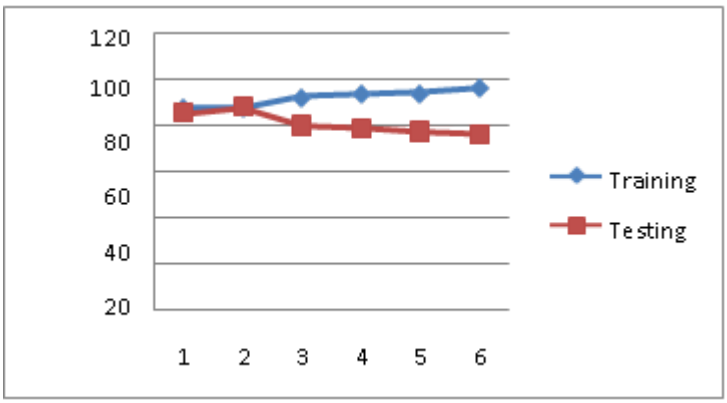
For CC view images, as a first attempt, data is analyzed on the basis of set 23 textural features obtained by merging the parameters selected by above mentioned three feature selection approaches, extracted from ROI (8x8) under μ±3σ normalization. Data clustering is verified in *NDA* projection space under different architectural settings of *ANN* classifier, as presented in Table 5.4. Then classifier is trained for all the settings to which best clustering of data is observed. Classifier testing and training performance under different number of neurons is summarized in Graph 6.1.




**Graph 6.1:Training and testing of *ANN* with different number of neurons under μ±3σnormalization for ROI (8x8).**

In this case we received the best classifier testing accuracy of 91.18% at the learning rateη=0.15 with 2 neurons in hidden layer of classifier.

To have better results, we tried to analyze the same data by applying 1-99% normalization.Similar to previous attempt features were extracted from ROI (8x8) and same method of features selection and verification of data clustering is adopted. Here, again the classifier is trained and tested for the settings for which we achieved the best clustering (results are presented in Table 5.6);performance of the classifier for this approach is summarized in Graph 6.2.



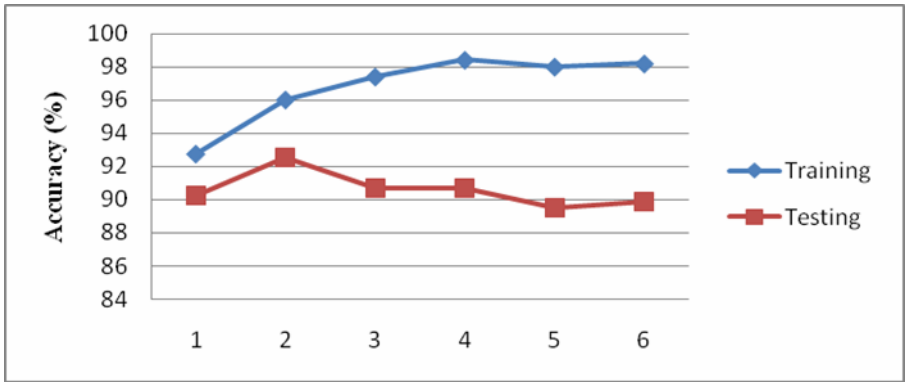

**Graph 6.2:Training and testing of *ANN* with different number of neurons under 1-99%normalization for ROI (8x8).**

Again, in this case we received an accuracy of 88.44% in testing phase when the classifier is

configured with2 neurons in hidden layer and a learning rate of 0.15, which is less than the previous case when data is normalized byμ±3σ approach for the same size of ROI (8x8).

As it is clear from the results presented above, the output performance of the system is not satisfactory, so, to have better results the data is classified by same procedures, but, the features are extracted on the basis of ROIs with window size (16x16), as recommended bya number of researchers (discussed in Section 5.4), under both normalization conditions.

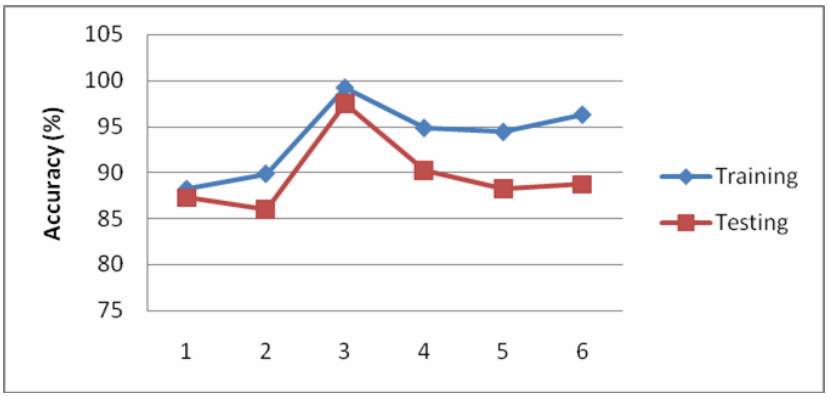
Again, in first attempt data is normalized by 1-99% approach and features are extracted from ROIs (16x16). When the classifier is trained tested for all the configurations for which we received the best clustering of data in *NDA* projection space (presented in Table 5.8), system performance is summarized in Graph 6.3.




**Graph 6.3:Training and testing of *ANN* with different number of neurons under μ±3σnormalization for ROI (16x16).**

In this attempt a maximum classifier testing accuracy of 92.56% is obtained when the classifier was configured with a learning of 0.20 and 2 neurons in hidden layer.

In the last attempt, CC view mammograms data is analyzed on the basis of parameters extracted from ROI (16x16) under μ±3σ normalization condition. The most relevant features are selected by hybrid feature selection approach, and then data clustering is verified in *NDA* projection space under different architectural settings, results are presented in Table 5.10. For the configuration to which we obtained best clustering, the classifier is trained and tested, the results are summarized in Graph 6.4.




**Graph 6.4:Training and testing of *ANN* with different number of neurons under μ±3σnormalization for ROI (16x16).**

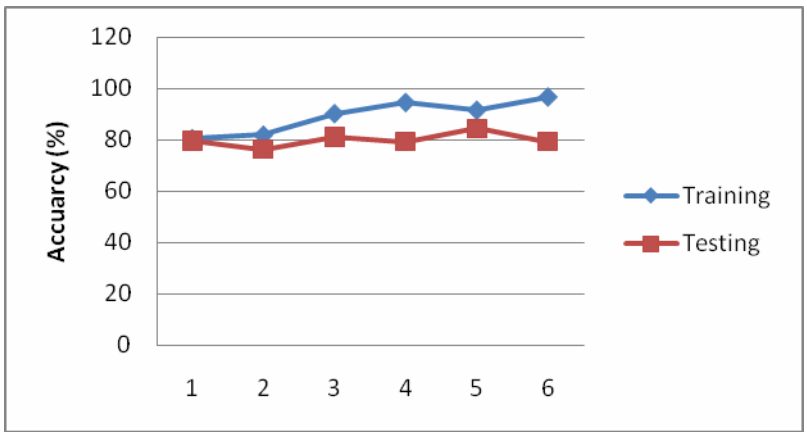
We achieved an excellent result with an accuracy of 99.21% and 97.55% in training and testing phase, when the classifier is configured with a learning rate of 0.15 and 3 neuronsin hidden layer.

It can be concluded from the results discussed here that for CC view mammograms our proposed system has shown the best performance when data is normalized by μ±3σapproach and features are extracted from ROI (16x16).

Because, a large number of abnormal regions in mammograms, as marked by radiologists, are small in size, so, ROIs having window size larger then (16x16) are not tried here.

Literature survey revealed that all researchers have analyzed CC and MLO view images,by machine vision approach all together in their work, but in this research, we have under taken MLO view images

separately in Section 5.5. The data is analyzed following thesame procedures as have been adopted for CC view images. We received an accuracy of 84.41% in testing phase for which training accuracy of classifier is 91.59%, when the 5 neurons are employed with a learning rate of 0.22.




**Graph 6.5:Training and testing of *ANN* with different number of neurons under μ±3σnormalization for ROI (16x16) for MLO view data.**

The results presented above show that the system performance is poor for MLO view images as compared to CC view images. As, it has been already discussed in Section 2.2 that both types of images (CC, and MLO) are acquired at right angle to each other, due to which pectoral muscle position and orientation, as well as, pixels distribution in both types of images is entirely different to each other, it may be one of the reasons due to which the system performance is lower for MLO view images.

We received the best results for ROI (64 64)test data images, but, for experimental data (digital mammograms) ROI with window size (16x16) has shown the best results. The reason behind this is that the test images were not true fine and micro, due to large kernel size, while for experimental data images, this condition of fine and micro-texture is truly fulfilled.

Similarly to some other researchers (discussed in Section 4.1.2), for mammograms we received better results under μ±3σ normalization conditions, but for test data (wheat kernelimages) we received good results for 1-99% approach. It may also be due to the difference in textures in both types of images.

**Conclusion**

In this study, our main purpose is to differentiate the abnormalities present in mammograms, as benign and malignant, via texture analysis. For this purpose 77 statistical textural features are calculated from ROI (8x8)and (16x16), under two normalizationconditions, μ±3σ and 1-99%, by using MaZda software and the best descriptors are achieved by a hybrid model of feature selection, based on combined set of features obtained by merging the features selected by three approaches, *F, POE+ACC* and *MI.* On the basis of these selected features data analysis is performed by three multivariate techniques *PCA, LDA,* and *NDA,* and the best performance is received by *NDA* approach which based on *ANN* classifier.

Under different architectural settings, based on neurons in input layer and learning rate „η" , *ANN* classifier is trained and tested by using two disjoint sets of data with the ratio 70% and 30% respectively. For ROI (8x8)developed in CC images under both normalization, μ±3σ and 1-99% we received the best results, with an accuracy of 91.18% and 88.44% in testing phase respectively, when the classifier is configured by 2 neurons in hidden layer and η=0.15. In the case of ROI (16x16) for normalization 1-99% we received an accuracy of 92.56 with the same configuration, but for μ±3σ our system has shown the best performance with an accuracy of 97.55%, when 3 neurons are employed in hidden layerbut again learning rate has the same default value (η=0.15).

From this experimental work, it is concluded that for CC view images, *ANN* classifier with 3 neurons hidden layer at learning rate 0.15 is the best configuration to differentiate abnormalities as benign and malignant.

For MLO view images, the said classifier has shown the best result with an accuracy of 84.41%, for 5 neurons and at learning rate 0.22, which is very poor result as compared to CC view images.

**Future work**

It must be noted that the ultimate goal of automated texture analysis in medical imaging is to improve the diagnostic accuracy in differentiation of the benign and malignant lesions and diseases staging in types of organs by providing a computer output as a second opinioninstead of pathological biopsy. So, future work can include:

1. Study the performance system for images of other modalities like, MRI, UT, EIT, *etc*.,
2. Study the performance of this CAD system for the remaining architectural factors *i.e.,* back propagation iterations and optimization iterations, because, in this work only neurons and learning factors are considered.
3. Search out the set of parameters to have better results for MLO images, because, onthe basis implemented parameters, we received poor performance; as compared to CC images.
4. Observer performance studies that (1) estimate improvement of radiologists‟ diagnostic accuracy in distinguishing task and (2) estimate the benefit of our automated distinguishing scheme in clinical management.

## References


[1] National Cancer Instiute, What You Need To Know About Breast Cancer, U.S.N.I.O. Health, Editor. 2000, Nih Publicaton No. 00-1556: Washington

[2] Loukas, C., Et Al., Breast Cancer Characterization Based On Image ClassificationOf Tissue Sections Visualized Under Low Magnification, In Computer And Mathematical Methods In Medicine. 2013, Hindawi Publishing Co. P. 1-7.

[3] Breast Cancer Prevention And Detection; Facts And Figures 2003-2004, In American Cancer Society. 2004: Atlanta, Ga.

[4] Breast Cancer Facts And Figures 2009-2010, In American Cancer Society 2010:Atlanta, Ga.

[5] Breast Cancer Facts And Figures 2012, The American Cancer Society: Ga, Usa.

[6] Ferlay, J., Et Al., Estimates Of The Cancer Incidence Mortality In Europe 2006. Annals Of Oncology, 2007. 18(5): P. 581-592.

[7] Levi, F., Et Al., Monitoring The Decrease In Breast Cancer Mortality In Europe. European Journal Of Cancer Prevention 2005. 14(6): P. 497-502.

[8] Khokher, S., Et Al., Knowledge, Attitude And Perventive Practices Of Women For Breast Cancer In The Educational Institutions Of Lahore, Pakistan. Asian Pecific Journal Of Cancer Pervention, 2011. 12: P. 2419-2423.

[9] El-Azim, M.H., N.S. El-Nagger, And H.Y. Syed, Knowledge, Attitude And Practices Regarding Breast Self-Examination Among Female UndergraduateStudents In The Faculity Of Applied Medical Sciences At Umm Al-Qura University Ksa Journal Of American Sciences, 2013. 9(4): P. 622-632.

[10] Lim, G.C.C., S. Rampal, And Y. Halimah, Cancer Incidence In Peninsular Malaysia, 2007-2008, In National Cancer Registry. 2008: Kuala Lumpur.

[11] Sobani, Z.A., Et Al., Knowledge, Attitude And Practices Among Urban Women Of Karachi, Pakistan, Regarding To Breast Cancer. Journal Of Pak. Med. Assoc., Nov.,2012. 62(11): P. 1259-1264.

[12] Malik, I.A., Clinico-Pathological Features Of Breast Cancer In Pakistan Journal Of Pak. Med. Assoc., 2002. 52: P. 100-104.

[13] Howlader, N., A.M. Noone, And M. Krapcho, Seer Cancer Statistics Review 1975-2008. 2008, National Cancer Institute: Md, Usa.

[14] Gomez, S.S., Et Al., Impact Of A Cad System In A Screen-Film Mammography Screening Program: A Prospective Study. European Journal Of Radiology, 2011. 80: P. 317-321.

[15] Thamgavel, K. Mammogram Image Segmentation Using Fuzzy Clustering. In Ieee International Conference On Pattern Recognition, Informatic And Medical Engineering 2012.

[16] Colak, S.B., Et Al., Clinical Optical Tomography And Nir Spectroscopy For Breast Cancer Detection. Ieee Journal Of Selected Topics In Quantum Eelectronics, 1999. 5: P. 1143-1158.

[17] Haaggesen, C.D., Disease On The Breast. W. B. Saunders, Company. 1971.

[18] Benaron, D.A. And D.K. Stevenson, Optical Time-Of-Flight And Absobance Imaging Of Biologic Media. Science, 1993. 259: P. 1463-1466.

[19] Khuzi, A.M., Et Al., Identification Of Masses In Digital Mammograms Using Gray Level Co-Occurance Matrices. Journal Of Biomedical Imaging And Intervention, 2009. 5(3): P. 1-13.

[20] Buseman, S., Et Al., Mammography Screening Matters `For Young Women With Breast Carcinoma. Cancer, 2003. 97(1): P. 352-358.

[21] Heinlein, P., Ed. Robust Techniques For Enhanncement Of Microcalcifications In Digital Mammography. Medical Imaging System Technology, Ed. C.T. Leondes. Vol. A-5. 2005, World Scientific Publishing Co. Ltd.: Singapore.

[22] Kallergi, M., Computer-Aided Diagnosis Of Mammographic Microcalcificaion Clusters. Medical Physics, 2004. 31(2): P. 314-326.

[23] Sickles, E.A., Mammographic Features Of 300 Consective Non-Palpable Breast Cancers. American Journal Of Roentgenology 1986. 146: P. 661-663.

[24] Highman, R.P. And M. Brady, Mammographic Image Analysis. 1999, Doodrecht: Kluwer Academic Publishers.

[25] .Dengler, J., S. Behrens, And J.F. Desaga, Segmentation Of Microcalcifications InMammograms. Ieee Transactions Of Medical Imaging, 1993. 12(4): P. 634-642.

[26] Diyana, W.M., J. Larcher, And R. Baser, Comparisom O ClusteredMicrocalcifications Automated Detection Methods In Digital Mammograms, In Ieee International Conference On Accoustic, Speech And Signal Processing 2003. P. 385-388.

[27] Bozek, J., K. Delac, And M. Grgic. Computer-Aided Detection And Diagnosis Of Breast Abnormalities In Digital Mammography.In 50th International Symposium, Elmar-2008. 2008. Zadar, Croatia.

[28] Bird, R.E., T.W. Wallace, And B.C. Yankaskas, Analysis Of Cancers Missed At Screening Mammography. Radiology, 1992. 184: P. 613-617.

[29] Cheng, H.D., Et Al., Computer-Aided Detection And Classification Of Microcalcifications In Mammograms: A Survey Pattern Recognition, 2003. 36: P. 2967-2991.

[30] Timp, S. And N. Karssemeijer, A New 2d Segmentation Method Based On Dynamic Programming Applied To Computer-Aided Detection In Mammography MedicalPhysics, 2004. 31(5): P. 958-971.

[31] Warren, L., S. Wood, And C. Diorsi, Potential Contribution Of Computer-Aided Detection In The Sensitivity Of Screening

Mammography. Radiology, 2000. 215: P. 554-562.
[32] Giger, M. And H. Macmohan, Image Processing And Computer-Aided Diagnosis. Radiol. Clin. North America, 1996. 24(3): P. 565-596.
[33] Elmore, J.G., Et Al., Variability In Radiologists' Interpretation Of Mammograms. N. Eng. J. Med., 1994. 331: P. 1493-1499.
[34] El Naqa, I., Ed. Techniques In The Detection Of Microcalcification Clusters InDigital Mammograms. Medical Imaging Systems Technology, Ed. C.T. Leondes. Vol. A 5. 2005, World Scientific Publishing, Co. Ltd.: 5 Toh Tuck Link, Singapore.
[35] Juntu, J., Et Al., Machine Learning Study Of Several Classifiers Trained With Texture Analysis Features To Differentiate Banign From Malignant Soft TissueTumors In T1-Mri Images. Journal Of Magnetic Resonance Imaging 2010. 31: P. 680-689.
[36] Lee, J.K.T., Interpretation Accuracy And Pertinence American College OfRadiology, 2007. 4: P. 162-165.
[37] Mousa, R., Q. Munib, And A. Mousa, Breast Cancer Diagnosis System Based On Wavelet Analysis And Fuzzy-Neural Expert Systems With Applications, 2005. 28(4):P. 713-723.
[38] Al-Hadidi, B., M.H. Zu'bi, And N.H. Suleiman, Breadt Cancer Image Detection Using Image Processing Functions Journal Of Informaton Technology, 2007. 6(2): P. 217-221.
[39] Huo, Z., Et Al., Effectiveness Of Cad In The Diagnosis Of Breast Cancer: An Observer Study On An Independent Database Of Mammograms. Radiology, 2000. 7:P. 1077-1084.
[40] Chan, H.P., Et Al., Improvement Of Radiologists' Characterization Of Mammographic Masses By Using Computer-Aided Diagnosis An Roc Study. Radiology, 1999. 212: P. 817-827.
[41] Cernadas, E., Et Al., Eds. Design Of Unsharp Masking Filters In The Frequency Domain: Parameterization For Breast Radiography Digital Mammography, Ed. K. Doi, M. Giger, And R.M. Nishikawa. 1996, Elsevier, Amsterdam.
[42] Soares, F., Et Al. In Ieee Second International Conference On System, Networking And Communication. 2007.
[43] Gonzalez, R.C. And R. Woods, Digital Image Processing. 3rd Ed. 2008, Upper Sadle River Pearson Printice Hall.
[44] Julesz, B., E. Gilbert, And L.A. Sheep, Inability Of Human To Dscriminate Between Visual Textures That Agree In Second-Order Statistics, 1973. 2: P. 391-405.
[45] Julesz, B., Experiments In Visual Perception Of Textures. Scientific American,1975. 232(4): P. 34-43.
[46] Belotti, R., Et Al., A Completely Automated Cad System For Mass Detection In Large Mammographic Database. Medical Physics, 2006. 33(8): P. 3066-3075.
[47] Makinaci, M., Support Vector Machine Approach For Classification Of Cancerous Prostate Regions. World Academy Of Science, Engineering And Technology, 2005. 7: P. 166-169.
[48] Jirari, M., Cad System For Digital Mammograms Based On Radial Basis Functions And Feature Extraction Techniques In Proceedings Of The 27th Ieee EngineeringIn Medicine And Biology 2005: Shanghai, China. P. 4457-4460.
[49] Haralick, R.M., K. Shanmugam, And I. Distein, Texture Features For Image Classification. Ieee Transactions On System, Man And Cybernetics, 1973. Smc-3:P. 610-621.
[50] Dash, M. And H. Liu, Feature Selection For Classification. Interantional Journal Of Intelligent Data Analysis, 1997. 1: P. 131-156.
[51] Szczypiński, P.M., Et Al., Mazda—A Software Package For Image Texture Analysis. Computer Methods And Programs In Biomedicine, 2009. 94(1): P. 66-76.
[52] Heinlein, P., Ed. Robust Techniques For Enhancement Of Microcalcifications In Digital Mammograms`. Medical Imaging Systems Technology, Ed. C.T. Leondes. 2005, World Scientific Pub. Co. Ltd.: 5 Toh Tuck Link Singapore.
[53] Bassett, L.W., Et Al., Diagnosis Of The Diseases Of The Breast. 1997, Singapre: W.B. Saunders Company.
[54] Http://Www.Webmd.Com/Women/Picture-Of-The-Breasts.
[55] Chu, K.C., C.R. Smart, And R.E. Tarone, Analysis Of Breast Cancer Mortality And Stage Distribution By Age For Health Insurance Plan Clinical Trial. Journal Of National Cacer Institute, 1988. 14: P. 1125-1132.
[56] Boyle, P. And B. Levin, World Cancer Report 2008, . 2008, International Agency For Research On Cancer: Lyon, France.
[57] Kopans, D.B., Breast Imaging. 2nd Ed. 1998, Philadelphia, Ny: Lippincott-Raven Publishers.
[58] Singh, R.T., T.S. Wadhawani, And S. Singh, 3rd Uksim European Symposium Modeling And Simulation. Ieee Computer Society, 2009. Ems2009: P. 251-256.