Chapter 2
DIGITAL MAMMOGRAPHY

Early detection of breast cancer has great significance in reducing mortality. Identifying the breast cancer at an early stage using CAD systems is especially to help radiologists to make their decisions more accurate with minimum time and cost. This chapter focuses on the different aspects of the breast cancer as well as the main objectives of designing an effective CAD system for early detection of breast cancer. The chapter begins with a discussion of the breast anatomy followed by a concise description of the stages and types of breast cancers. The processes of screening as well as the different imaging modalities are introduced. Finally an overview of CAD system highlighting the significant steps and parameters for evaluating the performance of CAD system is also presented.
2.1 Structure and Function of Breast

The important anatomical structure of a breast is shown in Figure 2.1.

Breasts are made up mostly of fat and breast tissue, along with nerves, veins, arteries and connective tissue that helps hold everything in place. The main chest muscle (the pectoral muscle) is found between the breast and the ribs in the chest wall. Breast tissue is a complex network of lobules (small round sacs that produce milk) and ducts (canals that carry milk from the lobules to the nipple openings during breast feeding) in a pattern that looks like bunches of grapes. These “bunches” are called lobes. Adult women have 15 to 20 lobes in each breast. Each lobe has 20 to 40 lobules. Small ducts are attached to the lobules. These ducts join together like branches of grape stems.
into increasingly larger ducts. There are about ten duct systems in each breast, each with its own opening at the nipple.

Muscle tissue in the nipples allows them to become erect in response to stimulation or breast feeding. Muscle tissue around the lobules helps squeeze milk into the ducts. Glands on the areola (the shaded circle of skin around the nipple) secrete small amounts of fluid to lubricate the nipple when breastfeeding [http://www.komen.org, 2008]

The entire structure of breast can be classified into two components. The first component is concerned with milk production known as the epithelial components. The second component consists of fat and connective tissue, which supports and protects the structure of the breast.

The epithelial component of the breast consists of a tree-like branching pattern of milk ducts that come together at the nipple. The leaves of this tree are formed by the lobules which are the secretory units of the breast. Each lobule consists of a number of acini connecting to an intra-lobular duct. The acini are composed of two types of cells, namely, the epithelial and myo-epithelial. The epithelial cells secrete variety of glyco-proteins and during lactation they also produce milk. The myo-epithelial cells are capable of contracting during breast feeding. Each intra-lobular duct connects with an extra-lobular duct, and this together with the lobule, is called the terminal ductal lobular unit [Nagi, 2011].

The extra-lobular ducts of the breast link together and form sub-segmental ducts, which in turn form the segmental ducts. These ducts drain milk from different segments or lobes of the breast. In total, the breast consists of 15 to 20 lobes, which are roughly pyramidal in shape with the apex directed towards the nipple. The non-epithelial component of the breast
consists mainly of fatty tissue. There are no muscles in the actual breast, but there are a series of muscles behind and underneath the breasts. These muscles work together with a ligament called Cooper ligament to support the weight of the breast [Nagi, 2011].

Breasts contain lymph vessels, which are very important in fighting against diseases in the body. Lymph is a clear fluid that contains tissue fluid, waste products and immune system cells. The lymph system consists of lymph nodes and lymph vessels that transport the lymph to the lymph nodes. Most of the lymph vessels that go through the breast carry the lymph to the lymph node underneath the arm pit, called auxiliary nodes. The other lymph vessels carry the lymph to the lymph nodes which are inside the chest, called the internal mammary nodes, or the lymph nodes above or below the collarbone, called the supraclavicular or infraclavicular nodes respectively. Lymph veins can also carry possible diseases to lymph nodes which might increase the spread of the disease, for example breast cancer (malignant) cells [Boyle and Levin, 2008] [Nagi, 2011].

Breast cancer is developed when the cells of the breast become abnormal (malignant) and spread without order or control. The malignant cells then form a tissue and turn into a tumor. This tumor typically grows into nearby tissues or breaks away and enters the bloodstream or lymphatic system which can affect other organs. The spreading of breast cancer is generally referred to as metastasis [Nagi, 2011]. The most common and effective method for detecting breast tumors in their early stages is by performing mammogram screening. Mammography is currently the most effective modality used to detect tumors in the breast tissue that can indicate potential clinical problems, such as the: asymmetries between breasts, architectural distortion, confluent densities associated with benign fibrosis,
microcalcification clusters (MCCs) and mass lesions. The two most common features that are typically associated with breast tumors are MCCs and mass lesions.

### 2.1.1 Calcification

![Figure 2.2: Microcalcification clusters (MCC) in a breast tissue](image)

Breast calcifications are tiny calcium deposits that develop in the breast tissue as woman ages. They are common and are usually benign (noncancerous) found on about half of all the mammograms. Calcifications can result from a number of different things such as inflammation and trauma to the breast. There are two types of breast calcifications: Macro calcification and microcalcification. Macrocalcifications are look like large white dots on a mammogram and are often dispersed randomly within in the breast. These dots are usually associated with benign conditions and rarely require breast biopsy. Macrocalcifications are common and they are found in approximately half of women over age 50 and one in 10 women under age 50. It is considered as noncancerous. Micorcalcifications on the other side are small calcium deposits that look like white specks on a mammogram. Individual microcalcifications typically range in size from 0.1 to 1.0mm with an average diameter of about 0.5mm. Microcalcifications are usually not a result of cancer. But if they appear in certain patterns as clustered together, that may
be a sign of precancerous cells or early breast cancer. A Radiologist carefully studies the shape and size of the calcium deposits to determine if they require further study. Most often, microcalcifications have nothing to do with cancer. However, in about 30% of cases they can be related to an early breast cancer. [Nagi, 2011][Kneece, 2008] [http://ww5.komen.org, 2008].

2.1.2 Mass Lesion

Breast tumor is often represented as a mass lesion with or without the presence of MCCs. A cyst, which is a non-cancerous collection of fluid, may appear as a mass in the film. However, ultrasound or fine needle aspirations can distinguish the difference. The similarity in intensities with the normal tissue and morphology with other normal textures in the breast makes it more difficult to detect masses compared with calcifications [Feig and Yaffe, 1995]. The location, size, shape, density, and margins of lesions are useful for the radiologist in evaluating the likelihood of a cancer [Evans, 1995]. Most benign masses are well circumscribed, compact, and roughly circular or elliptical, as shown in Figure 2.3. Malignant lesions usually have a blurred boundary, irregular appearance and sometimes are surrounded by a radiating
pattern of linear spicules [Evans, 1995]. However, some benign lesions may have a speculated appearance or blurred periphery.

2.2 Breast Tumor

Breast tumor is an uncontrolled growth of breast cells. It occurs as a result of mutations, or abnormal changes, in the genes responsible for regulating the growth of cells and keeping them healthy. The genes are in each cell’s nucleus, which acts as the “control room” of each cell. Normally, the cells in our bodies replace themselves through an orderly process of cell growth: healthy new cells take over as old ones die out. But over time, mutations can “turn on” certain genes and “turn off” others in a cell. That changed cell gains the ability to keep dividing without control or order, producing more cells just like it and forming a tumor [www.breastcancer.org, 2012].

Unfortunately, Scientists are not sure about the exact cause of breast cancer, but they have identified high risk factors for this disease. The most common factors include: age, family history and personal history. The most common symptom is a painless lump in the breast. At times, a painful lump in the breast turns out to be cancer (malignant). One or more lumps in the armpit can be a symptom of breast cancer; however, they can also be due to non-cancerous (benign) conditions. The bleeding from the nipple can indicate the presence of cancer, especially if the bleeding occurs from one breast only [Buseman et.al, 2003]. A more difficult symptom to be identified is the thickening of the tissue in the breast. Any changes in the breast size or shape can be due to cancerous (malignant) or non-cancerous (benign) conditions. Although benign conditions in the breast can be understood by their symptoms, biopsies need to be performed to understand whether the
irregularity in the breast is non-cancerous or otherwise. Other symptoms indicating the possibility of breast cancer are: redness of the skin over a portion of the breast, redness or scaliness of the nipple or any nipple pain or retraction (nipple turning inward), an orange peel appearing on the skin and dimpling of the skin [Bassett et al, 1997].

The term *breast tumor* refers to a malignant tumor that has developed from cells in the breast. Usually breast tumor either begins in the cells of the lobules, which are the milk-producing glands, or the ducts, the passages that drain milk from the lobules to the nipple. Over time, cancer cells can invade nearby healthy breast tissue and make their way into the underarm lymph nodes, small organs that filter out foreign substances in the body. If cancer cells get into the lymph nodes, they then have a pathway into other parts of the body. The breast cancer’s stage refers to how far the cancer cells have spread beyond the original tumor. The cancerous breast tumor (malignant) can be classified into two categories: (i) Non-Invasive breast cancer and (ii) Invasive breast cancer. The following section explains about these two types of cancers.

### 2.2.1 Non-Invasive Breast Cancer

Non-Invasive-in situ-cancer consists of malignant cells that replace the normal epithelial cells, lining the ducts or lobules in the breast tissue. The Non-Invasive means it would not spread out of ducts or lobules where the cancer was originated. The two forms of Non-Invasive cancer are: (i) Ductal Carcinoma In Situ (DCIS) and (ii) Lobular Carcinoma In Situ (LCIS).

- Ductal Carcinoma In Situ (DCIS) is the most common type of non-invasive breast cancer. Ductal means that the cancer starts inside the milk ducts, carcinoma refers to any cancer that begins in the skin or
other tissues (including breast tissue) that cover or line the internal organs, and in situ means "in its original place." DCIS is called non-invasive because it hasn't spread beyond the milk duct into any normal surrounding breast tissue. DCIS is not a life-threatening, but it has the possibility of increase in the risk of developing an invasive breast cancer later. It is estimated that about one-third to half of the untreated patients eventually will develop invasive cancer, usually in the same quadrant of the breast where the first lesion develops. Mammographically DCIS is often characterized by the presence of microcalcifications. When there is extensive fibrosis, DCIS may also present as a palpable mass\cite{Lu and Bottema, 2001}\cite{www.breastcancer.org, 2012}

- Lobular carcinoma in situ (LCIS) is an area (or areas) of abnormal cell growth that increases a person's risk of developing invasive breast cancer later on in life. Lobular means that the abnormal cells start growing in the lobules, the milk-producing glands at the end of breast ducts. Carcinoma refers to any cancer that begins in the skin or other tissues that cover internal organs—such as breast tissue. In situ or “in its original place” means that the abnormal growth remains inside the lobule and does not spread to surrounding tissues. LCIS rarely gives rise to mammographic abnormalities. It is often found in biopsies that have been done for other reasons such as removal of benign lesions. LCIS is a risk factor for developing breast cancer. The majority of patients are therefore managed by careful follow ups \cite{Lu and Bottema, 2001}. 

2.2.2 Invasive Breast Cancer

Invasive breast cancer, also known as *infiltrating cancer*, occurs when malignant cells have spread beyond the ducts or lobules to other parts of the breast or body. Invasive cancers vary in size from less than 10mm in diameter to over 80mm, but are usually 20 to 30mm at presentation [Vitak, 1998]. There are two forms of Invasive cancer: (i) Invasive Ductal Carcinoma (IDC) and (ii) Invasive Lobular Carcinoma (ILC).

- **Invasive ductal carcinoma (IDC),** sometimes called infiltrating ductal carcinoma, is the most common type of breast cancer. About 80% of all breast cancers are invasive ductal carcinomas. *Invasive* means that the cancer has “invaded” or spread to the surrounding breast tissues. *Ductal* means that the cancer began in the milk ducts, which are the “pipes” that carry milk from the milk-producing lobules to the nipple. *Carcinoma* refers to any cancer that begins in the skin or other tissues that cover internal organs such as breast tissue. All together, “invasive ductal carcinoma” refers to cancer that has broken through the wall of the milk duct and begun to invade the tissues of the breast. Over time, invasive ductal carcinoma can spread to the lymph nodes and possibly to other areas of the body.

- **Invasive lobular carcinoma (ILC),** sometimes called infiltrating lobular carcinoma, is the second most common type of breast cancer after invasive ductal carcinoma (cancer that begins in the milk-carrying ducts and spreads beyond it). According to the American Cancer Society, in each year more than 180,000 women in the United States find out that they have invasive breast cancer. About 10% of all invasive breast cancers are invasive lobular carcinomas.
2.2.3 Different Types of Tumors

Even though, there are many different types of breast abnormalities; it is possible to have a general classification as either benign or malignant type. The basic differences between the benign and malignant types are normally identified by using the boundary shapes with surrounding breast tissues. This differentiation can be performed by examining speculations on the malignant tumor that can be easily identified using mammography or ultrasound techniques.

Speculation is a satellite distortion caused by the intrusion of breast cancer into the surrounding tissue and its existence is very important for tagging the tumor as malignant. Breast tumors and masses appear as dense regions in mammography. A typical benign mass has a round, smooth, and well-circumscribed boundary whereas a malignant mass has a speculated, rough and blurry boundary as shown in below Figure [Varela et al, 2006] [Cheng et al, 2006].

![Benign and Malignant Tumors](image-url)

**Figure 2.4:** Benign and malignant tumors. (a and b analyzed by [Rangayyan et.al, 1997] c and d [Guliato et.al, 2006])
As far as Radiologist is concerned, it is very difficult to differentiate between benign and malignant tumors on mammograms. But many recent studies have shown that techniques can be developed to assist radiologists to decide on the type of tumor by using quantitative method. [Guliato et. al, 2006] derived a mathematical model for mammograms, in order to derive polygonal models of contours for an accurate classification of tumors. [Rangayyan et.al, 2000] developed a method to quantify the sharpness of the tumor boundaries. [Varela et. al, 2006] divided the tissue tumor border into three sections and analyzed these sections independently to decide if the mass lesion is benign or malignant by considering the shape of its interface. [Kim and Min, 2002] developed a mathematical model to count the number of jags of the breast tissue and breast tumor interface in order to differentiate between benign and malignant tumors. The most common and effective method for detecting breast tumors in their early stages is by performing mammogram screening.

2.3 Screening for Breast Cancer

Early detection of breast cancer can save thousands of woman lives each year. Screening is a very important step for early detection of breast cancer which locates breast cancers since they are still small in size and confined to the breast before they cause any symptoms. In many countries breast cancer screening programs using mammography have been started to detect cancers as early as possible. A screening program is defined as a program where an asymptotic group is invited to examine a specific disease on a regular basis. For breast cancer screening, a number of parameters must be chosen for the purpose. Out of which two parameters are most significant. They are: (i) the age range of women that are invited and, (ii) the time interval between two screening rounds. Below the age of 40, the incidence
rate of breast cancer is extremely small, increasing rapidly between the age of 40 and 50, which continues to increase more gradually for older women. The problem with screening young women is that their breasts contain much glandular tissue, yielding mammograms that are difficult to read due to denser tissues [Nagi, 2011].

Breast cancers in young women are often aggressive and fast growing tumors, requiring short intervals between two screenings. After menopause, the breast becomes less dense, making successful screening for small cancers more feasible. The upper limit of age for which women are invited for screening varies between 65 and 75 [Van Dijck et.al, 1997]. If the interval between two successive screening rounds is too large, a number of tumors that are detected in screening have already reached a stage with a lower chance of successful treatment. Tumors occur during this interval are known as interval carcinomas. A large number of interval carcinomas may indicate that the screening interval should be made shorter. A short interval period will have a larger effect on the reduction of mortality, but is more expensive and women are exposed to a higher number of X-ray doses. In the United Kingdom, the screening interval is 3 years, a period that is considered too long by some researchers [Dean, 1996], whereas in Sweden and the Netherlands it is 2 years [Tabar et.al, 1987]. It is found that breast cancer is the commonest cancer found among Malaysian women which is 18 % of all cancers [Lim et.al, 2008]. However, in Malaysia, there is no national screening program for breast cancer.

The way by which mammograms are read also varies between countries. In some countries (for example the Netherlands) mammograms are examined by two radiologists, called double reading. Various approaches can be used to combine the findings of the two radiologists. [Thurfjell et.al, 1994]
found that the sensitivity increases when double reading is practiced (when a case is recalled if either one of the radiologists finds it suspicious), without changing the positive predictive value (PPV). In medical diagnostic tests using ROC curves, sensitivity represents the ratio of tumors which are marked and classified as tumor, to all marked tumors. Specificity represents the ratio of tumors which are not marked and also not classified as tumor.

2.4 Imaging Modalities

At present digital mammography is considered as the best imaging modality available for breast cancer screening. In addition to digital mammography, other modalities like ultrasonography (US) and Magnetic Resonance Imaging (MRI) are also used for further analysis, when digital mammography is not sufficient. The following section gives an overview of these two imaging modalities used for mammography screening.

2.4.1 Ultrasonography

Ultrasound is non-ionizing, real time, portable and inexpensive compared with other clinical imaging modalities. However, images can be difficult to interpret, requiring expert training. In addition, organs such as the brain located beneath bone cannot be imaged clearly. Nevertheless, ultrasound is particularly functional for obstetrics (fetal imaging) and quantification of blood flow using Doppler measurements [Feng, 2008]. The role of Ultrasonography (US) in breast imaging is a subject of ongoing discussion. Studies that have been performed using US as a mammogram screening tool failed to establish its efficiency. Thus, it has been concluded that US should not be used as a mammogram screening tool [Nagi, 2011] [Rahbar et.al, 1999].
2.4.2 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-ionizing technique with excellent soft-tissue contrast and high spatial resolution. The temporal resolution is typically much lower than the ultrasound with scans lasting several minutes. The cost of MRI scanners is relatively high, and the large superconducting magnet requires special housing in clinical environments. The major uses of MRI are in the areas of brain disease, spinal disorders, angiography, cardiac assessment and musculoskeletal damages [Feng, 2008].

High resolution Magnetic Resonance Imaging (MRI) of the breast has recently emerged as a sensitive instrument for the detection of breast cancer. MRI has proven to be useful in screening younger women with dense breasts who are at a high risk of developing breast cancer [Stoutjesdijk et. al, 2001]. MRI can also be used as an adjunct to mammography for selected patients. However, MRI has a significant false positive (FP) rate and it is not available in all areas due to being more expensive than digital mammography. Other limitations of MRI are that, it requires contrast injection and it can cause problems with claustrophobia. Thus, at the moment MRI remains limited to specific problem solving situations for patients at high risk for cancer.

2.4.3 Mammography

Mammography is acknowledged as the single most effective method of screening for breast cancer. Mammography based screening programs are carried out in many countries, and their effectiveness has had a great impact on the diagnosis. It is the one of the significant modality in medical image acquisition mechanism for the early detection of the breast cancer. Here we discuss on the development of a CAD system based on image processing
techniques for the automated detection and classification of breast tumor using digital mammograms [Feig, 1988].

During the last two decades, CAD systems have been developed to help the radiologists in the interpretation of mammograms but leaving the final decision to the radiologist. The goal of this technique is the detection of the disease at a pre-symptomatic phase. When the symptoms are developed, the cancer has typically become invasive and consequently the prognosis is less favourable [Oliver et.al 2010]. The CAD scheme identifies small regions of potential clustered microcalcifications on the digital mammograms. Generally it is desirable to improve the sensitivity (the fraction of the cancer cases that is correctly detected as cancers) of the CAD scheme in order to detect the most subtle cases. However, as the sensitivity increases with the current CAD scheme, the false-positive detection rate (the fraction of the normal cases that is incorrectly detected as cancers) will also increase [Yoshida et.al, 1994].

Mammograms can detect tumors that are an eighth of an inch in diameter, while manual examination usually fails to detect tumors smaller than a half-inch. Screening for breast disease via mammography depends on observing local and distant changes in tissues. Important visual clues of breast cancer include preliminary signs of masses and calcification clusters [McLelland, 1990][ Chang and Laine, 1999]. Unfortunately, at the early stages of breast cancer, these signs are very subtle and varied in appearance, making diagnosis difficult, challenging even for specialists [Kegelmeyer et.al, 1994]. Radiographic signs of cancer are related to tumor mass and density, size, shape, smoothness of borders and calcification distributions. Extraction of these features and enhancement of them assist radiologist to locate suspicious areas more reliably. [Xing et.al, 1995]
Due to the low contrast of the mammogram images or high frequency content of the noise, it is very hard to see the small breast cancer lesions in the mammography [Mayo et.al, 2004], it can greatly reduce the breast cancer mortality in a well organized screening program over the population. The performance of the mammography decreases as the density of the breast increases. This situation is inconvenient since breast cancer risk increases as the breast density increases [Oliver et.al, 2010].

2.5 Identifying Breast Changes in Screening

Screening is for identifying the cancer before a person has a feel of any symptoms. This will help to find cancer at an early stage. When cancer is found early, it is easier to treat. By the time symptoms appear, cancer may have begun to spread. Three tests are commonly used to screen for breast cancer.

- Mammogram analysis: Taking the X-ray of the breast.
- Clinical Breast Examination (CBE): A clinical breast exam is an examination of the breast by a doctor or other health professional. The doctor will carefully feel the breasts and under the arms for lumps or anything else that seems unusual.
- Breast Self Examination (BSE): Breast self–examination refers to examination to check their own breasts for lumps or anything else that seems unusual.

While screening mammography attempts to identify breast cancer in the asymptomatic population, diagnostic mammography are performed to further evaluate abnormalities such as palpable mass in a breast or suspicious findings identified by screening mammography.
Breast Positioning in Digital Mammography

In digital mammography, the breast is compressed between two parallel plates to spread the breast tissue and make the breast a block of uniform physical thickness for the X-rays to pass through. This compression can be performed at different angles to generate different orientations of the breast. The two standard views used in screening mammography are the: Cranio-Caudal (CC) view, generating a top to bottom view of the breast and the Mediolateral Oblique (MLO) view, a side-on view at approximately 45°. Examples of each view are shown in Figure 2.5 with the CC view shown in Figure 2.5(a) and MLO view in Figure 2.5(b).

While the breast is compressed to a uniform physical thickness during mammography, the radiographic density of each tissue type present in the breast determines the appearance of the mammogram. Radiographic density is the term used to describe the level of attenuation that the X-rays experience from the source to the detector. The higher the density, the less developed the film, resulting in appearance from fully exposed (black) to unexposed (white) depending on the tissue type. The fat in breasts has a low density, allowing the X-rays to pass through easily to expose the film, hence fatty areas of the mammogram are dark, in some places almost as dark as portions of the image where there is no tissue (background pixels). The glandular tissue in the breast has a somewhat higher density, resulting in brighter areas, as does the tissue of the pectoral muscle. Microcalcifications are very high in density; some lesions also have high density.
2.7 Interpretation of Mammogram Images by Radiologists

The American College of Radiology (ACR) developed a standard mammographic interpretation system called Breast Imaging Reporting and Data System (BI-RADS) lexicon for describing the characteristics of the abnormality including the final pre-pathology findings for the Radiologists. This system uses shape, borders and relative intensity features for the classification of mass lesions in an image. BI-RADS descriptors are important factors for predicting malignancies that are assessed and provided by the radiologists. Mass narratives include the overall shape description, the border region margin regularity and the relative intensity of the mass region compared with the ambient normal tissue intensity. The BI-RADS lexicon provides a four category rating for assessing the overall breast tissue characteristics in terms of fibro-glandular composition. The composition categories relate to the degree of interpretation difficulty. Similarly, the BI-RADS gives a five-point overall assessment that is related to the degree of probable malignancy. BI-RADS have established mass descriptors such as...
for shape and margins for the detection of mass lesions as indicated in below Figure. The shape and margin properties of BI-RADS descriptors are as follows:

- **Shape**: The shape of the mass is described with a five-point assessment: round, oval, lobular, irregular and architecturally distorted as shown in Figure 2.6(a).

- **Margin**: The mass margins modify the boundaries. For example the overall shape of the mass may be round, but close inspection may reveal scalloping along the border, which may indicate a degree of irregularity or a lobular characteristic. The margins are rated with a five-point system: circumscribed (well-defined/sharply-defined) margins, obscured margins, micro-lobulated margins, ill-defined margins and speculated margins as shown in Figure 2.6(b).

### 2.8 Computer-Aided Diagnosis Systems

There is a substantial research regarding the detection and classification of masses and calcifications in breast. These problems are generally considered well studied, and new developments must meet or exceed the high standards of performance set by the existing algorithms. Moreover, commercial CAD systems have satisfactory effectiveness detecting masses and calcifications. Even so, certain areas of research in CAD of breast cancer still require attention [Rangayyan et.al, 2007][Cruz, 2011]. The common CAD systems include image acquisition, Image enhancement, Segmentation or Extraction of region of interest of the enhanced image. After extracting the region of interest using segmentation technique, the objects or regions characteristics are identified and labelled using feature extraction techniques. These features should have similar values.
for objects in the same categories and different ones for different categories in order to distinguish them and such features should be selected for further analysis. Finally the selected features are used for the classification [Sampat et.al, 2005]. The Figure 2.7 depicts the various processes involved in Computer Aided diagnosis system.

![Figure 2.6: BI-RADS mass descriptors for (a) shape (b) margin](image)

Mammography lesions such as microcalcifications and masses are usually small and have low contrast regarding to the contiguous breast tissues, thus these abnormalities are hard to detect. Image enhancement can improve the radiologists’ perception to subtle diagnosis and consequently to more accurate diagnosis [Rangayyan et.al, 2007][Cruz,2011]. Image enhancement techniques include contrast and intensity manipulation, additional reduction of noise, background removal, edge sharpening and filtering. The usual task of mammogram enhancement is to increase the contrast between regions of interest (ROI) and background and to sharpen the edges or borders of regions of interest [Cheng et al, 2003].
After the enhancement of the mammogram, the next step is the segmentation which extracts the objects or relevant regions for further examination. It reduces the time and size of the data for processing. The segmentation can be done in order to obtain locations of suspicious areas to assist radiologists for diagnosis or to classify the abnormalities as benign and malignant [Cheng et.al, 2003]. There are numerous image segmentation techniques available for extracting regions of interest in mammogram images.

In the next step, which is the most important as well as crucial one, we extract the most relevant features in the mammogram images for identifying and classifying the images. Different feature extraction techniques and feature selection methods exists for retrieving potentially useful features for classification.

Finally the classification techniques such as KNN, Lazy and ANN classifiers are used for classifying the images based on selected features. There are two different strategies for classifying individual regions in medical images. They are region-based and pixel-based classification. In region-based classification scheme, an object is first segmented from the
image. The object is then classified based on object level features computed over an entire object. Normally the extracted features from region or object are shape, texture, intensity, size, position etc. within the image. On the other hand, in pixel-based classification, feature value is computed for each pixel. Each pixel in the image is then classified individually. A pixel level feature is considered as a feature which is computed in the neighbourhood of a pixel and is associated with that pixel [Meyer-Base, 2004].

2.9 Performance Evaluation of CAD systems

In order to evaluate the performance of any CAD system which performs pattern classification (developed using supervised machine learning techniques such as ANNs and SVMs), the binary classification performance has to be measured. The two important evaluation techniques for evaluating the classification performance in such problems are the confusion matrix and ROC curves. In this thesis we make use of the confusion matrix obtained during the classification as the performance evaluation criteria.

2.9.1 Confusion Matrix

A convenient tool when analyzing results of classifier systems in general is a confusion matrix, which is matrix containing information about the actual and predicted classes. The matrix is a two dimensional and has as many rows and columns as there are classes. The column represents the true classification and the rows represent the system classification. If the system performs perfectly, there will be scores only in diagonal positions. If the system has any misclassifications, they are placed in the off-diagonal cells [Meyer-Base, 2004] [Han et al, 2012]
### Table 2.1: A confusion matrix showing relation between TP, FP, FN and TN

<table>
<thead>
<tr>
<th>Confusion Matrix</th>
<th>Positive (p⁺)</th>
<th>Negative (n⁻)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (p⁺)</td>
<td>True Positive (TP)</td>
<td>False Negative (FN)</td>
</tr>
<tr>
<td>Negative (n⁻)</td>
<td>False Positive (FP)</td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

#### 2.9.2 Sensitivity, Specificity and Accuracy

The performance of a binary classifier can be best described in terms of its sensitivity and specificity, quantifying its performance to false positive (FP) and false negative (FN) instances. It can be then evaluated based on the data available in the confusion matrix [Sakka et.al, 2005] [Han et.al, 2012]. An image or Region of Interest (ROI) may be classified as either cancerous (positive) or normal (negative). The final decisions belong to any four possible categories: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). FN and FP represents two kinds of errors. An FN error implies that true abnormality was not detected and a FP error occurs when a normal region was falsely identified as abnormal image. A TP decision is correct judgment of an existing abnormality and a TN decision means that a normal region was correctly labeled [Sakka et. al, 2005][Khuzi et.al, 2009]. Therefore the accuracy and performance of any CAD system is evaluated based on the Sensitivity, Specificity and Accuracy. They are defined as follows:

Sensitivity is defined as the ratio of tumors which are marked and classified as tumor, to all marked tumors, given by:

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2.1)
\]

Specificity is defined as the ratio of tumors which are not marked and also not classified as tumor, to all unmarked tumors, given by:
Specificity = \frac{TN}{TN + FP} \quad (2.2)

The overall accuracy is the ratio between the total number of correctly classified instances and the test set size, given by:

\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.3)

### 2.10 Summary

In this chapter we presented a detailed account of digital mammography and its application for the early detection of breast cancer. Chapter begun with the description of anatomical structure and function of breast followed by a brief summary of breast tumor and the screening program available for identifying the tumor at the very beginning stage. Interpretation of mammogram images using BI-RADs description is also explained. Finally a detailed description of the computer aided diagnosis scheme and its performance evaluation criteria are given.

\ldots...\ldots...