Early detection of breast cancer mass lesions by mammogram segmentation images based on texture features

Faleh H. Mahmood

Remote Sensing Unit, College of Science, Baghdad University, Baghdad, Iraq
Email: faleh_sine@yahoo.com

Abstract
Mammography is at present one of the available method for early detection of masses or abnormalities which is related to breast cancer. The most common abnormalities that may indicate breast cancer are masses and calcifications. The challenge lies in early and accurate detection to overcome the development of breast cancer that affects more and more women throughout the world. Breast cancer is diagnosed at advanced stages with the help of the digital mammogram images. Masses appear in a mammogram as fine, granular clusters, which are often difficult to identify in a raw mammogram. The incidence of breast cancer in women has increased significantly in recent years.

This paper proposes a computer aided diagnostic system for the extraction of features like mass lesions in mammograms for early detection of breast cancer. The proposed technique is based on a four-step procedure: (a) the preprocessing of the image is done, (b) regions of interest (ROI) specification, (c) supervised segmentation method includes two stages performed using the minimum distance (MD) criterion, and (d) feature extraction based on Gray level Co-occurrence matrices (GLCM) for the identification of mass lesions. The method suggested for the detection of mass lesions from mammogram image segmentation and analysis was tested over several images taken from Al-Iliwiya Hospital in Baghdad, Iraq. The proposed technique shows better results.

Keywords
Breast Cancer, Mammogram, Masses, calcifications, segmentation, Co-occurrence matrices.

Article info
Received: Mar. 2012
Accepted: May. 2012
Published: May. 2012
Introduction

Cancer is a group of diseases that cause cells in the body to change and grow out of control. Most types of cancer cells eventually form a lump or masses called a tumor, and are named after the part of the body where the tumor originates. Breast cancer begins in breast tissue, which is made up of glands for milk production, called lobules, and the ducts that connect lobules to the nipple. The remainder of the breast is made up of fatty, connective, and lymphatic tissue[1]. Breast cancer is a leading cause of cancer deaths among women. Efficient detection is the most effective way to reduce mortality, and currently a screening programme based on mammography is considered one the best and popular method for detection of breast cancer. Mammography is a low-dose x-ray procedure that allows visualization of the internal structure of the breast. Mammography is highly accurate, but like most medical tests, it is not perfect. On average, mammography will detect about 80%-90% of the breast cancers in women without symptoms. Testing is somewhat more accurate in postmenopausal than in premenopausal women [2].

Breast image analysis can be performed using mammography, magnetic resonance, nuclear medicine or ultrasound. So far the most effective and economical breast imaging modality has been mammography due to its simplicity, portability and cost effectiveness. Digital mammography is a technique for recording x-ray images in computer code instead of on x-ray film, as with conventional mammography. Digital mammography may have some advantages over conventional mammography. The images can be stored and retrieved electronically. Despite these benefits, studies have not yet shown that digital mammography is more effective in finding cancer than conventional mammography [3]. The images are displayed on a computer monitor and can be enhanced (lightened or darkened) before they are printed on film. Images can also be manipulated; the radiologist can magnify or zoom in on an area. From the patient’s perspective, the procedure for a mammogram with a digital system is the same as for conventional mammography[4]. Initial mammographic or MRI images themselves are not usually enough to determine the existence of a benign or malignant disease with certainty. If a finding or spot seems suspicious, your radiologist may recommend further diagnostic studies. Interpretations of mammograms can be difficult because a normal breast can appear differently for each woman. Recent studies showed that the interpretation of the mammogram by the radiologists give high rates of false positive cases. Several research works have tried to develop computer aided diagnosis tools. They could help the radiologists in the interpretation of the mammograms and could be useful for an accurate diagnosis [5]. Imaging techniques play an important role on mammogram images, especially of abnormal areas that cannot be physically felt but can be seen or processed on a conventional mammogram or with ultrasound. In this paper we have proposed a new technique, and we have developed a supporting tool for easy identification of abnormal masses in mammography images.

The paper is organized as follows: Section 2 presents the brief description of mammographic abnormalities. Section 3, presents the brief description of image analyzing systems Section 4 presents the feature extraction, Section 5 presents data sources Section 6 detailed the proposed methodology for mass lesion detection Section 7 describes results and discussion. Conclusion is summed up in Section 8.
X-Ray Mammography

X-Ray Mammography is commonly used in clinical practice for diagnostic and screening purposes. Screening mammography has been recommended as the most effective method for early detection of breast cancer. Mammography provides high sensitivity on fatty breast and excellent demonstration of micro calcifications; it is highly indicative of an early malignancy. Due to its low cost, it is suitable for mass screening program. Mammography has its limitations. It is less reliable on dense breast of young women or women underwent a surgical intervention in the breast because glandular and scar tissues are as radiopaque as abnormalities. Furthermore, there is low dose X-Ray radiation [6].

1. Mammographic Abnormalities

Numerous characteristics that may signify a probable clinical problem, including asymmetries between the breasts, architectural distortion, and confluent densities associated with benign fibrosis, calcifications and masses and the like are identified with the aid of mammography. The two most customary characteristics associated with cancer are clusters of masses and calcifications both of which are discussed subsequently. The detection of calcification has been explored by various groups of researchers. Small (sub 15mm), low contrast masses are considered more critical than micro-calcifications, since they are more difficult to detect [7]. Of chief concern are the masses that are not accompanied by micro calcifications since they are tumors that develop drastically. Unlike micro calcifications that are well apparent as bright spots, the masses merge with the breast structure in such a way that boundaries are indistinct, and can often be completely hidden from vision if the breast is dense [8].

2. Mass Lesions

Breast cancer is characterized with the presence of a mass accompanied or not accompanied by calcifications. There is a possibility of a cyst, which is non-cancerous collection of fluid to resemble a mass in the film. The identical intensities of the masses and the normal tissue and similar morphology of the masses and regular breast textures makes it a tedious task to detect masses in comparison with that of calcifications. The location, size, shape, density, and margins of the mass are highly beneficial for the radiologist to evaluate the probability of cancer. A majority of the benign masses are well circumscribed, compact, and roughly circular or elliptical whereas the malignant lesions are characterized by blurred boundaries, irregular appearances and are occasionally enclosed within a radiating pattern of linear spicules, see Fig.(1). Nevertheless some benign lesions may also possess speculated appearances or blurred peripheries [7].

3. Calcification

Tiny deposits of minerals (calcium) that appear like localized high-intensity regions (spots) in the mammogram are known as calcifications. Calcifications are one of the significant and widespread finding that are frequently apparent in a mammogram. Micro-calcifications and macro calcifications or coarse calcifications are the two common categories of calcifications. Macro-calcifications are coarse calcium deposits that are spread about the breast. Commonly such deposits are accompanied by benign conditions and hardly necessitate a biopsy. The benignity or malignancy of the tumor is indicated by the number calcifications that comprise a cluster.
Micro-calcifications are minute (less than 1/50 of an inch or ½ of a millimeter) spots of Calcium deposits that may exist in an area of rapidly dividing cells. They are possibly intramammary, inside and around the ducts, inside the lobules, in vascular structures, in interlobular connective tissue or fat. The onset of cancer might be indicated by the presence of micro-calcifications in a cluster. Almost half of the cancers identified through mammography come into sight as cluster of micro-calcifications. Generally ductal carcinoma in situ (an early cancer confined to the breast ducts) is identified when micro-calcifications become apparent through mammography. The morphology of calcifications is considered to be the most important indicator in differentiating benign from malignant [7]. Three categories of calcifications have been identified by the “The American College of Radiology (ACR) BIRADS” (Table 1):
(a) Typically benign
(b) Intermediate concern
(c) High probability of malignancy

**Table (1) Summary of BI-RADS categories of calcifications [7]**

<table>
<thead>
<tr>
<th>Type of calcification</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skin</td>
<td>Typical lucent center and polygonal shape</td>
</tr>
<tr>
<td>Vascular</td>
<td>Parallel tracks or linear tubular calcifications that run along a blood vessel</td>
</tr>
<tr>
<td>Coarse or pop-corn like</td>
<td>Involuting fibroadenomas</td>
</tr>
<tr>
<td>Rod-shaped</td>
<td>Large rod-like structures usually &gt; 1mm</td>
</tr>
<tr>
<td>Round</td>
<td>Smooth, round clusters</td>
</tr>
<tr>
<td>Punctate</td>
<td>Round or oval calcifications</td>
</tr>
<tr>
<td>Spherical or lucent centered</td>
<td>Found in debris collected in ducts, in areas of fat necrosis</td>
</tr>
<tr>
<td>Rim or egg-shell</td>
<td>Found in wall of cysts</td>
</tr>
<tr>
<td>Milk or calcium</td>
<td>Calcium precipitates</td>
</tr>
<tr>
<td>Dystrophic</td>
<td>Irregular in shape but usually large &gt; 0.5mm in size</td>
</tr>
</tbody>
</table>

### Image Analysis Systems

Generally, image analyzing systems involve the operations illustrated in Fig.(2). The Preprocessing improves the quality of the data by reducing the existed artifacts, while the Feature Extraction and Selection provides the measurement to facilitate the image segmentation process. Finally, image Segmentation and Classification group pixels into regions and define the boundaries of the various regions [9].

**Fig. 1. (a) Dense breast containing a malignant mass. (b) Fatty and glandular breast containing a malignant mass[7].**
Image segmentation is a crucial step toward image interpretation, since the rest of the analysis fully relies on the data from this step. Subsequent, medical processing and analysis steps may include quantification, registration, visualization and computer aided diagnosis. Therefore, image segmentation is one of the most important tasks in computer vision and image processing.

Micro-calcifications and macro-calcifications or coarse calcifications are the two common categories of calcifications. Macro-calcifications are coarse calcium deposits that are spread about the breast. Commonly such deposits are accompanied by benign conditions and hardly necessitate a biopsy. The benignity or malignancy of the tumor is indicated by the number calcifications that comprise a cluster [10], it can be performed either manually or using certain image processing and computer vision techniques. Generally, image segmentation techniques locate objects consisting pixels having something in common; e.g. having similar intensity values or same colors [11].

In this paper, both statistical and structural features will be used to segmenting normal or/and abnormal brain’s tissues.

Feature Extraction
The transformation of an image into set of features is known as feature extraction. In the literature, different approaches have been proposed to extract the suitable sets of features; e.g. statistical-based approach, structural-based approach, model-based approach, and transform-based approach [12]. In this paper, we shall concentrate on the statistical-based approach, using the 1st and 2nd order statistics, to identify abnormalities on tissues of breast’s mammogram images.

1. Statistical Features Analysis
A statistical feature is one of the well known and simplest methods, so far used, to measure the image texture behaviors. Normally, image analysis systems used two methods for extracting the image texture features; i.e. utilizing either 1st order histogram features, the 2nd order of the co-occurrence matrix features, or utilizing higher order of invariant moments features [13].

1.1 First-Order Statistics Features
Histogram summarizes the statistical information about the image. For an image $f(x, y)$ of dimensions $N\times M$ and $G$-gray values, its histogram (referred as probability density function “pdf”) is presented by, [14];

$$\rho(z_i) = \frac{n(z_i)}{M \times N} , \ z_i = 0,1,\ldots\ldots,G-1 \quad - - - - (1)$$

The probability function $\rho(z_i)$ of occurrence of the intensity level $z_i$ is obtained by dividing the number of the intensity level $n(z_i)$ by the total number of pixels in the image.

Quantitatively, there are certain useful features can be obtained from the image’s histogram [15]; e.g.

**a. The image Mean value:**

$$\mu = \sum_{i=0}^{G-1} z_i \rho(z_i) \quad (2)$$

**b. The image Variance value:**

$$\sigma^2 = \sum_{i=0}^{G-1} (z_i - \mu)^2 \rho(i) \quad (3)$$
c. The image Entropy value:

\[ H = - \sum_{i=0}^{G-1} \rho(z_i) \log \rho(z_i) \]  

1.2 Second - Order Statistics Co-Occurrence Features.

It has been shown above, the 1st order histogram based features are local in nature; i.e. they haven’t reflected any spatial information. For this purpose, the gray-level spatial co-occurrence matrices have been designed to represent 2nd order histogram features based on the joint probability distribution of pairs of pixels.

Haralick [16] suggested the use of gray level co-occurrence matrices (GLCM) for definition of textural features. The values of the co-occurrence matrix elements present relative frequencies with which two neighboring pixels separated by distance “d” appear on the image at an angle “θ” (usually \( d = 1,2, \ldots \), and \( \theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ \)). As an example, let a pair of pixels \( f(i, j) \) and \( f(m, n) \) in an image, separated by distance “d” at an angle “θ” with respect to the horizontal axis, see Fig. (3) . Such matrix is symmetric and also a function of the angular relationship between two neighboring pixels. The co-occurrences matrix can be calculated on the whole image, but by calculating it in a small window which scanning the image, the co-occurrence matrix can be associated with each pixel. By using gray level co-occurrence matrix we can extract different features like probability, entropy, energy, variance, inverse moment difference etc.[17]. Three of these 2nd order measures have been adopted in this research and used to differentiate between breast’s tissues, i.e.

\[ \text{INR} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - j)^2 \rho_{r,0}(i, j) \]  

\[ \text{LOCH} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\rho_{r,0}(i, j)}{1 + (i - j)^2} \]  

\[ \text{Cor} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\rho_{r,0}(i, j)}{\sigma_x \sigma_y} \]  

Where: \( \mu_x, \mu_y, \sigma_x, \sigma_y \) are the rows and columns means and standard deviations

Data Sources

The proposed research makes use of the data collection obtained from AL-ILWIYA Hospital in Baghdad, Iraq. The collection comprises of 50 images that fall into one of the following categories: normal, benign and malign. Malign images are regarded as abnormal. Additionally, the malign cases are further classified into namely: circumscribed
masses, speculated masses, ill-defined masses, architectural distortion and asymmetry. Every image is digitized at a resolution of 1024x1024 and 512x512 pixels and 8-bit accuracy (gray level).

The Proposed Technique
In this paper, the supervised segmentation technique using minimum distance algorithm was implemented, certain areas should be selected, referred as region of interest “ROIs”, from which statistical features are counted and saved for further processing. The input image then scanned top to bottom, using certain window’s size, the same statistical features previously counted for the ROI should be computed for each window’s position. The window’s features should then be compared with those of the ROIs, the window’s center point assign a class number matched with nearest ROI features. Matching process is usually performed either by implementing the Minimum Distance (MD). In our present research, the minimum distance measure is preferred because of its simplicity and requiring less computation time.

Our proposed technique was based on a four-steps procedure: (a) the preprocessing of the image is done b) regions of interest (ROI) specification, (c) the two stages supervised segmentation method includes performed using the minimum distance MD criterion as a third step, and (d) feature extraction based on Gray level Co-occurrence matrices (GLCM ) for the identification of mass lesions. The method suggested for the detection of mass lesions from mammogram image segmentation and analysis was tested over several images semi-automatic supervised segmentation methods have been suggested and used for isolating the mass lesions from the healthy of breast tissues. The methods, as will be illustrated, are based on utilizing the first and higher orders statistical features.

1. Preprocessing
Mammograms are medical images that are difficult to interpret, thus a preprocessing phase is needed in order to improve the image quality and make the segmentation results more accurate. The first step involves the removal of unwanted parts in the background of the mammogram. The main objective of this process is to improve the quality of the image, to make it ready for further processing. Removing the irrelevant parts of the image is done by increasing contrast of the mammogram using threshold value.

2. Regions of Interest Specification
In order to segment the ROIs from breast tissue, it is presumed that pixels that constitute a ROI need to be members of a set of adjoining neighbor pixels with appropriate intensities. The “minimum intensity threshold” and “maximum intensity threshold” are the two thresholds that are utilized for determining the appropriate intensities. According to the observations the diameters of masses fall between upper and lower boundaries. Thus in order to comprehend if a pixel is present in the center region of the ROI, the diameter of the ROI needs to be considered foremost. The first step comprises of the manual choice of the number of ROIs for every mammogram. The location selections were made under the supervision of the radiologists involved in the study and this facilitated in obtaining non overlapping ROIs for every mammogram. The location selections were made under the supervision of the radiologists involved in the study and this facilitated in obtaining non overlapping ROIs for every mammogram, besides covering most of the breast density see Fig. (4). This location also guarantees that we carried out our assessment only with the breast tissue, devoid of the bias brought about by the pectoral muscle or imaging artifacts.
3. Two-Stage Supervised Segmentation Method

After selections different ROI’s from breast tissue, this supervised segmentation method includes two stages; in the first stage the 1st order statistical features (i.e. Mean, Variance, and Entropy) have been computed within certain defined window’s size. Since the extracted statistical features, in this stage, depends mainly on the image intensity values, the resulted segmented (or classified) image [see Fig.(5c)] hasn’t recognized the mass lesions from the healthy of breast tissues. To overcome this problem, a second stage segmentation operation is proposed to refining the result of the preliminary segmentation process. In this stage, only the mean values of the preliminary classes are computed (using the pixel’s values of the original image), then a reclassification operation is performed by assigning each classified point to its nearest mean value, illustrated in Fig.(5d). In fact, the second stage classification could be regarded as to be supervised classification process, because it has been performed on initially classified image points. It is remained to be noted that; the nearness measure between the counted means and the preliminary classified image pixels is performed using the minimum distance MD criterion. Results for two-stage supervised segmentation mammographic images are displayed in Fig. (5).

4. Mass Isolation Isolate using 2nd Order Statistical features

For more efficient result, the two-stage segmented image can be processed again by utilizing higher order statistical features. The co-occurrence matrices have been used to isolate the mass lesions from the surrounded contain of the breast. Different samples of mammogram images for normal and abnormal breast have been tested.

The refining operation by this method can be summarized by the following steps:

**Step1:** Input the original and the two-stage segmented image;

**Step2:** Click on original image to select number of points, as region of interest (ROI), based on final segmented image (represent the mass lesions features, having same classified color)

**Step3:** Define Window’s size and number of selected points in step2.

**Step4:** For each image window, compute the 2nd order co-occurrence features (i.e. Correlation, Local Homogeneity, and Inertia) (Equations. 5, 6 and 7);

**Step5:** Only those classified points having mass lesions color should be checked by comparing their Co-Occurrence features with those of the selected points in step4.

**Step6:** If the minimum distance (MD)between the window’s features and the closest features of the selected points was less than a decided threshold(Th), then the window’s center point is decided as to be mass lesions points; Otherwise.
Step 7: The window’s center is decided as to be non-mass lesions
Step 8: Continue till the end of the image
The results of above mentioned mass lesions isolation process for mammographic images as illustrated in Fig.(5). More results for another mammographic image as shown in Fig.(7).

Results and Discussion
As it has been mentioned in above, the adopted supervised segmentation the supervised segmentation technique requires to define priory; ROIs, number of classes, window’s size and minimum deviated error. Figures.(6-b and c) represent the results of different segmentation images obtained by 5×5 pixels window’s size, the Minimum Deviation Distance (MDD) equal 12, and for varying number of classes. The (MDD) has been chosen as to be "12" due to the gained results through our presence course of work. The minimum distance criterion is utilized to test the similarity or dissimilarity between breast components. Because of this technique is based on the intensity information in an image, for this reason, coloring overlapping was the problem faced us in isolating the mass area from the surrounded contain of the breast tissue. To improve the results and to avoid of this problem, an adaptive multi-stage segmentation by using the mean value of each pre-segmented classes has been introduced to extract the mass tissue from those showed similar behaviors. Then, the segmentation process is followed by extracting features using the Co-occurrence was regarded as 2nd order statistical features have been proposed and presented to differentiate affected breast tissue (i.e. Tumor) from other tissues. In our opinion that, the successive-stages of supervised segmentation method followed by extracted features 2nd order statistical features may be more successful for recognizing, identifying, and isolating tumor tissue in mammogram images.

Figure (5): Two-Stage supervised segmentation method, where a) preprocessing abnormal breast original image b) Segmentation Process c)Preliminary stage segmentation d) 2nd stage segmentation e) Final stage segmentation.
Figure (6): shown; (a) Original preprocessing breast tumor image, (b) Final stage segmentation (c) Image for first code-vector, (d) Image for second code-vector, (e) Image for third code-vector, (f) Image for fourth code-vector, (g) Image for fifth code-vector, (h) Image for sixth code-vector (mass lesions).

Figure (7): Refining results of all breast tissue classes using the co-occurrence features. (a) Original preprocessing breast tumor image, (b) Preliminary stage segmentation (c) Final stage segmentation (c) Image for mass lesions.
Conclusions

Breast cancer is one of the major causes of death among women. Digital mammography screening programs can enable early detection and diagnose of the breast cancer which reduces the mortality and increases the chances of complete recovery. Supervised successive segmentation method has been introduced to isolating the mass area from healthy of the breast tissues, based on certain 1st order desired features.

One of the main disadvantages acquired from the implementation of the supervised classification is that, generally requires a high level of expertise i.e. the region of interest “ROI” are delineated by some expert peoples (as experienced radiologists and physicians). The problems usually encountered when trying to design a semi-automatic system for tumor’s detection system is that; a large number of tumor’s types are existed, which differs in size, shape, location, tissue composition and homogeneity. But the main advantages acquired from the implementation of the supervised segmentation is that; numbers of regions are defined priory by the operators. Definitely, for abnormal breast, number of regions should be, at least, being increased by one of the normal brain tissues. Cases number of tissues class should be increased by two referring to the tumor and other healthy tissues.

The contemporary preference for the premature detection of breast cancer in women is Mammography. Nevertheless, the elucidation of mammograms greatly depends on radiologist’s opinion.

References