Automatic Segmentation and Identification of Abnormal Breast Region in Mammogram Images Based on Statistical Features

Faleh H. Mahmood*[1], Alaa Ali Hussein[2]
[1]Remote Sensing Unit, College of Science, University of Baghdad, Baghdad, Iraq
[2]Research and Developing Office, Ministry of Higher Education and Scientific Research, Baghdad.

Abstract

Breast cancer is one of the most common malignant diseases among women; Mammography is at present one of the available method for early detection of abnormalities which is related to breast cancer. There are different lesions that are breast cancer characteristic such as masses and calcifications which can be detected trough this technique. This paper proposes a computer aided diagnostic system for the extraction of features like masses and calcifications lesions in mammograms for early detection of breast cancer. The proposed technique is based on a two-step procedure: (a) unsupervised segmentation method includes two stages performed using the minimum distance (MD) criterion, (b) feature extraction based on Gray level Co-occurrence matrices GLCM for the identification of masses and calcifications lesions. The method suggested for the detection of abnormal lesions from mammogram image segmentation and analysis was tested over several images taken from National Center for Early Detection of cancer in Baghdad.

Key words: Mammogram, Breast Cancer, Masses, calcifications, segmentation, Region of Interest (ROI)

الانقسام التلقائي وتحديد المناطق غير الطبيعية للثدي في التصوير الشعاعي بالاعتماد على الخصائص الإحصائية

فالح حسن محمود*[١] علاء علي حسين[٢]
[١]جامعة بغداد، كلية العلوم، وحدة الاستشعار عن بعد، العراق، بغداد، [٢] دائرة البحث والتطوير، وزارة التعليم العالي والبحث العلمي، بغداد، العراق

الخلاصة

سرطان الثدي هو واحد من اكثر الامراض الشائعة الخبيثة بين النساء ، التصوير الشعاعي للثدي هو واحدة من الوسائل المتاحة في الوقت الحاضر للكشف المبكر عن سرطان الثدي خصوصا التشوهات غير الطبيعية، هناك اضرار مختلفة تميز سرطان الثدي كالتكتلات والتكلسات التي يمكن الكشف عنها بهذه التقنية .في هذا البحث تم اقتراح تقنية للكشف المبكر عن سرطان الثدي بمساعدة الحاسوب لغرض كشف واستخلاص التفاصيل مثل الاورام بهيئة تكتلات وتكلسات. وتتمثل التقنية او الطريقة المقترحة هذه على خطوتين: (أ) يتم تطبيق طريقة تقسيم غير المرشدة والتي تشتمل على مرحلتين يتم تنفيذها باستخدام معيار المسافة الصغرى (MD)، (ب) استخراج التفاصيل اعتمادا على الخصائص الاحصائية للصورة بالاعتماد على مصفوفة الحدوث (Gray Level Co-occurrence Matrices GLCM) لتحديد الاورام بهيئة تكتلات وتكلسات . تم اختبار الطريقة المقترحة للكشف عن الاضرار غير الطبيعية من تجزئة صورة الثدي، وتحليلها وقد تم اعتماد نماذج مختلفة من التصوير الشعاعي للثدي تم الحصول عليها من المركز الوطني للكشف المبكر عن السرطان في بغداد.

*Email: faleh_sine@yahoo.com

Introduction

Breast cancer becomes a significant health problem in the world. Early detection is the primary solution for improving breast cancer prognosis. Screening can be done through digital mammogram, ultrasound, magnetic resonance imaging (MRI) or breast Ultrasound produces a good contrast image but it does not contain enough detailed information which can be found in digital mammogram [1].

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 [2] .

Magnetic Resonance Imaging (MRI) is the most attractive alternative to Mammography for detecting some cancers which could be missed by mammography. In addition, MRI can help radiologists and other specialists determine how to treat breast cancer patients by identifying the stage of the disease [3].

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 tests 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 [4].

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.

Mammography Abnormalities

Numerous characteristics 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 [5]. 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 [6].

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 [5,7] . 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 figure1. Nevertheless some benign lesions may also possess spiculated appearances or blurred peripheries [5].

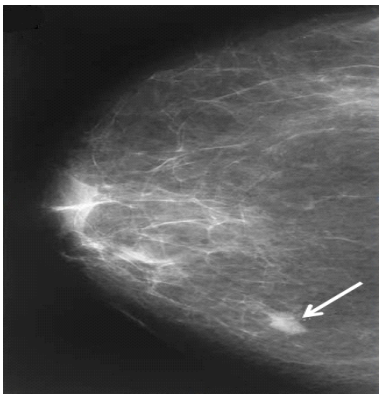


Figure1- Dense breast containing a malignant mass [8].

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 (an early cancer confined to the breast ducts) are 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 (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 [5].

Segmentation

In analyzing mammogram image, it is important to distinguish the suspicious region from its surroundings. The methods used to separate the region of interest from the background are usually referred as the segmentation [4]. Image segmentation is, by definition, the problem of decomposing images into regions that are semantically uniform. However, since images themselves provide only semantically poor information, image segmentation is essentially an application-oriented problem that demands either strong intervention of human experts or application specific solutions. Segmenting mammographic images into homogeneous texture regions representing disparate tissue types is often a useful preprocessing step in the computer-assisted detection of breast cancer [9].

The aim of the segmentation is to extract Regions of Interest (ROIs) containing all masses and locate the suspicious mass candidates from the ROI. Segmentation of the suspicious regions on a mammographic image is designed to have a very high sensitivity and a large number of false positives are acceptable since they are expected to be removed in later stage of the algorithm [10].

Statistical Features Analysis

Statistical texture features have been proven to be powerful in classifying masses and normal breast tissues. The implementation of feature extraction procedure relies on the quality of the texture, which is the main descriptor for all the mammograms. The gray level co-occurrence matrix (GLCM) is an advanced method of using different combinations of pixel brightness values which are the grey levels in the image. GLCM are constructed by observing pairs of image cells distance d from each other and incrementing the matrix position corresponding to the grey level of both cells [3]. 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. Table 1and 2 describes each of this features which are extracted from GLCM.

Table 1- 1st order statistical features [12].

Formula	Feature
$\mu = \sum_{i=0}^{G-1} z_i \rho(z_i)$	The image mean
$\sigma^2 = \sum_{i}^{G-1} (z_i - \mu)^2 \rho(i)$	The image Variance value
$H = -\sum_{z_i=0}^{G-1} \rho(z_i) \log \rho(z_i)$	The image Entropy value

Table 2- 2nd order statistical features[12].

Feature	Formula	Description
Correlation	$\sum_{i=0}^{c-1}\sum_{j=0}^{c-1}\frac{(i-\mu_x)(j-\mu_y)\rho(i,j)}{\sigma_x\sigma_y}$	represents the relationship between image's values
Homogeneity	$\sum_{i=0}^{c-1}\sum_{j=0}^{c-1}\frac{\rho(i,j)}{1+(i-j)^2}$	Measures the degree of homogeneity through image
Moment of Inertia	$\sum_{i=0}^{c-1}\sum_{j=0}^{c-1}(i-j)^2\rho(i,j)$	Measures the local contrast of an image

Where: μ_x, μ_y and *x,y* are the rows, columns mean and standard deviations respectively

Grey Level Co-Occurrence Matrices

The first order statistics of grey-level histograms give parameters for each processed region in the mammogram but do not provide any information about the repeating nature of texture. Co-occurrence matrices describe the level of gray level configuration and are used to explain the matrices of relative frequencies. Therefore the GLCM is an advanced method of using different combinations of pixel brightness values which are the grey levels in the image [13].

GLCM are constructed by observing pairs of image cells at distance d from each other and incrementing the matrix position corresponding to the grey level of both cells. This allows us to derive four matrices for each given distance and four different directions. The matrices are constructed at a distance of d = 1 pixel and for direction of θ given as 0°,45°,90° and 135° . A single direction might not give enough and reliable texture information. For this reason, four directions are used to extract the texture information for each masses and non-masses tiles area [14].

The Proposed Technique

1.Two-Stage Unsupervised Segmentation Method

In this paper, a semi-automatic unsupervised successive segmentation technique has been introduced to isolate abnormal breast tissue from healthy tissues, based on certain 1st order desired features. In our present research, the minimum distance measure is preferred because of its simplicity and requiring less computation time. The introduced technique includes two stages; firstly, for each image window, compute the 1st order statistical features (i.e. mean, standard deviation, and entropy).

Because of the unsupervised segmentation, the technique is based on the intensity information in an image. For this reason, coloring interference was the problem faced us in isolating abnormal area from healthy tissues. To overcome this color-overlapping problem, and to improve the result segmentation, an adaptation process as a second stage was proposed to refining the result of preliminary segmentation by assigning each classified point to its nearest mean value. The second stages has been done by computing the mean values of the preliminary classes the pixel's value of the raw image then reclassify the preliminary classified image according to the new mean values. The refining step is demonstrated as shown in figure. 2, and in table 3.

In fact, the second stage segmentation could be regarded as to be supervised segmentation, process, because it has been performed on initially segmented image points. It is remained to be noted that; the nearness measure between the counted means and the preliminary segmented image pixels is performed using the minimum distance (MD) criterion. The closeness between the counted means obtained by the supervised operation with the image pixel values is performed using the minimum distance creation method. Different samples of mammogram images for normal and abnormal breast have been tested, for different patients.

The following steps summarize the operations involved in the two-stage unsupervised segmentation algorithm:

Step1: Input mammogram Image for chosen patient.

Step2: Define Window's size (M), number of desired Classes and the Minimum Deviation

Distance (MDD) allowable deviated error between classes.

Step3: For each image window, compute the 1st order statistical features (i.e. mean, standard deviation, and Entropy see table 1.

Step4: The classification search is performed on the test image, pixel to pixel, from the upper left image corner; left to right, up to down.

Step5: A new set of features is counted, and compared with previous results which belong to past windows. For the first tested window, the pixel center classified as Class-1. Successively, the counted features are compared with those stored in feature vector to determine the closet class.

Step6: If the difference between the newly counted features with the stored was larger than a decided minimum deviation error, the pixel is assigned as new class number. Other it takes the class number belong to the closet absolute minimum features.

Step 7: If the number of the assigned class is found to be larger than the selected number of classes, then the pixel given a number out of the class's range, referring as to be undefined class. Thus, normally, numbers of classified regions are equal to the number of the decided classes plus one.

Step8: The preliminary segmentation process is terminated when the classification operation reach the last down-corner image point.

Step 9: Compute the mean values of each class.

Step 10: Reclassify the mammogram image, using the counted means (**step 9**), and stop. The results of above mentioned of two-stage unsupervised Segmentation process for mammographic images are illustrated in figure 2.

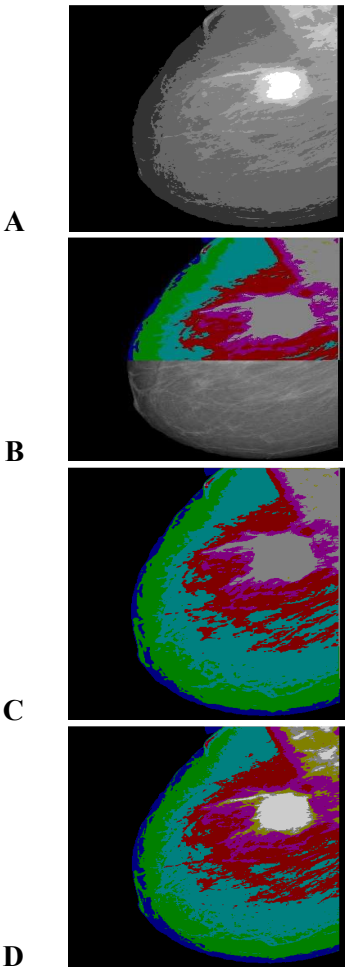


Figure 2- Two-Stage unsupervised segmentation method, where **a)** preprocessing abnormal breast original image **b)** Segmentation Process Steps **c)** Preliminary Segmentation **d)** Final Refining Segmentation

Table 3- Two-stage unsupervised segmentation method preliminary and final classified partitioning

Preliminary Classified Partitioning	Final Classified Partitioning
Ratio of Class No. 0 ٠,٣٤٨٩٨٨١١٨٣٧٢١٠٦	Ratio of Class No. 0 0.359988340458609
Ratio of Class No.1 ٣,٣٥٦٢٨٢٢١٩٤٤٩٢E-02	Ratio of Class No.) 3.5464433987563E-02
Ratio of Class No.2 ٠,١٢٣٨٨٩٥٦٧٤٨٦٥٣٦	Ratio of Class No. 2 0.146181500194326
Ratio of Class No.3 ٠,٣٢٩٨٤٠٠٩٧٧١٨٠٦١	Ratio of Class No. 3 0.201834989728499
Ratio of Class No.٤ ١,١٧٩٨٣٤٥٤٥٥٥٤٩E-04	Ratio of Class No.4 0.135340902781633
Ratio of Class No.5 ٥,٩٦٨٥٧٤٧٩٨٦٨٩٧E-04	Ratio of Class No. 5 7.1629373216368E-02
Ratio of Class No. ٦ ٠,١٦٨٣٢٧٦٨٨٦٣٤٧٢٣	Ratio of Class No. 6 5.4883126977579E-02

Total sum of Ratios	١,٠٠٥٣٢٣١٣٥٨٦١٤٢

2. Abnormal isolated regions based on 2nd order statistical features:

For robustness' purposes, the segmentation process is followed by a 2nd order classification method to assign image pixels accurately to their regions. The attached 2nd order method based on utilizing (Co-Occurrence matrices features). Different samples of mammogram images for normal and abnormal breast have been tested, for different patients. 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 represents the mass or calcification 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) see table 2.

Step5: Only those classified points having the mass or calcification lesions features 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.

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 abnormal isolation process for mammographic images as illustrated in figures 3 and 4 histogram of mammogram before and after segmentation.

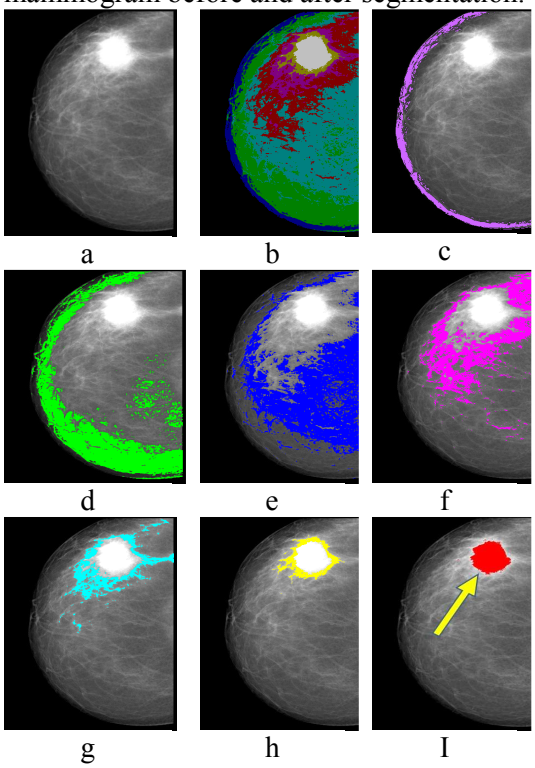
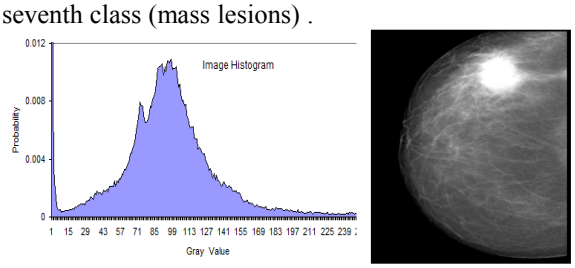
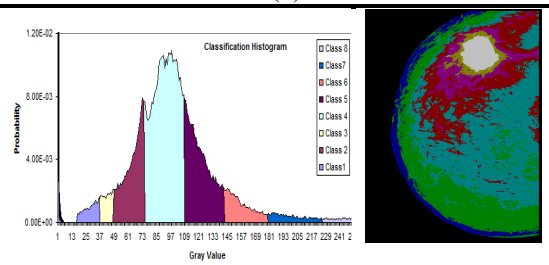


Figure 3- The (a) Original image, (b) Final stage segmentation (c) first class, (d) second class, (e) third class, (f) fourth class, (g) fifth class, (h) sixth class (i) seventh class (mass lesions) .



(a)



(b)

Figure 4- a) The mammogram grey scale image and its histogram b) The segmentation image and its histogram.

Results and Discussion

In this work, several normal and abnormal mammogram images have been used for ground truth purpose were randomly selected for texture extraction. It is found that the best features for discriminating masses are the features of GLCM

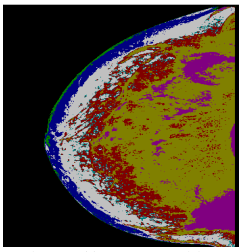
constructed at direction of 0°.

As it has been mentioned above, the adopted unsupervised segmentation method requires to define priory only number of classes, window's size and minimum deviated error. Figures 5-a and b present different segmented images obtained by 3×3 pixels window's size, the Minimum Deviation Distance **(**MDD**)** equal 12, and for varying number of classes. As can be seen, the preferable number of classes ranged between 7-10, which represent the main breast's tissues The method has been applied on both normal and abnormal breast images.

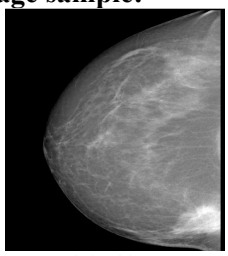
The (MDD) has been chosen as to be "12" due to the gained results through our presence course of work. The minimum distance criterion is utilized to test the similarity or dissimilarity between breast components. Because this technique is based on the intensity information in an image, coloring overlapping was the problem faced 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 was followed by extracting features using the Co-occurrence which was regarded as 2nd order statistical features.

This was proposed and presented to differentiate affected breast tissue (i.e. Tumor) from other tissues. In our opinion, the successive-stages of unsupervised segmentation method followed by extracted features 2nd order statistical features could be more successful for recognizing, identifying, and isolating tumor tissue in mammogram images.

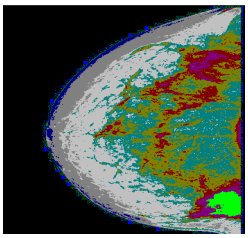
For abnormal image sample:



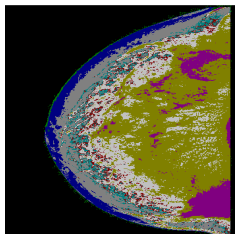
classes ٧Segmented by



Original image

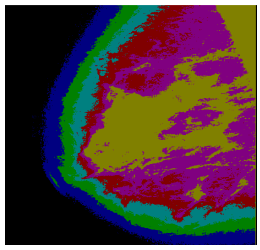


Segmented by 10 classes

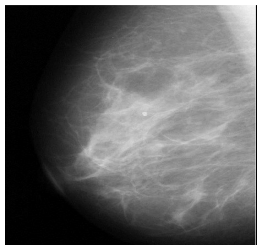


Segmented by 8 classes

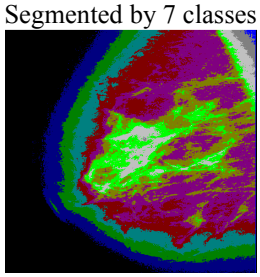
Figures 5- Examples of unsupervised successive segmentation results for mass, utilizing ROIs and different classes, using mask size=3×3 and MDD=12



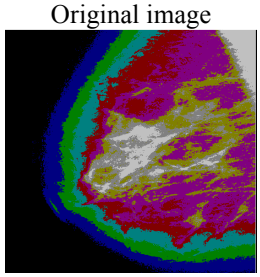
Segmented by 7 classes



Original image



Segmented by 10 classes



Segmented by 8 classes

Figures 6- Examples of unsupervised successive segmentation results calcification utilizing ROIs and different classes, using mask size = 3×3 and MDD = 12.

Conclusions

In this work, a semi-automatic unsupervised segmentation technique had been introduced to segment the breast tissues in a mammogram images. we have investigate and introduce a new method in an attempt to identify and isolate abnormal tissue (masses and calcification) from healthy breast tissues mammogram images, ranged from the easiest (i.e. 1st order statistical features e.g. mean, standard deviation and entropy) passing through 2nd order statistical features(e.g. Co-occurrence features).We Concluded that:

1- A better result has been obtained when we adopted the minimum distance (MD) criterion using unsupervised segmentation technique. The similarities and/or dissimilarities of the breast components became visible, by assigning different colors for different tissues.

2- Because the segmentation techniques are based on the intensity information and the inherent characteristics of mammogram images coloring overlapping was a problem that faced

us in isolating an abnormal (mass and calcification) area from a healthy breast tissue. In our opinion, the successive-stages of unsupervised segmentation method followed by extracted features 1st and/or 2nd order statistical features may be more successful for recognizing, identifying, and isolating abnormal tissue in mammogram images

3-In an attempt to identify and isolate abnormal tissue, the segmentation process is followed by extracted features using statistical features (mean and standard deviation as a 1st order and/or the Co-occurrence were regarded as 2nd order). Statistical features have been proposed and presented to differentiate affected breast tissue (i.e. mass or calcification) from other tissues.

4- The main advantage acquired from the implementation of the unsupervised segmentation is that the number of regions are defined prior by the operators. For abnormal breast tissue, number of regions should be, at least, increased by one class of the normal breast tissues. For best differentiation between normal and abnormal mammogram images, the results showed that the number of the decided classes should be 6 classes.

5-Definitely, for abnormal breast tissue, number of regions should be, at least, being increased by one of the normal breast tissues(i.e cases number of tissues class should be increased by one referring to the tumor and other healthy tissues) .

References

1. Samir K. B. **2010**. Pre-processing of Mammogram Images, *International Journal of Engineering Science and Technology*, 2(11), pp:6753-6758.
2. Saheb, B.,.K.Satya P. **2009**. Automatic detection of breast cancer mass in mammograms using morphological operators and fuzzy c-means clustering, *Journal of Theoretical and Information Technology*, 26(7), pp:704-709.
3. Nalini S.and Ambarish G M. **2011**. reast Cancer Mass Detection in Mammograms using K-means and Fuzzy C-means Clustering, *International Journal of Computer Applications*, 22(2), pp:15-21.
4. Indra K. M., Sanjay N.and Samir K. B. .**2011**. Identification of Abnormal Masses in Digital Mammography Images ,*International Journal of Computer Graphics*,.2(1), pp:17-29
5. Rabi N. P., Bijay K. P. and Manas R. P. **2009.** Feature Extraction for Classification of Microcalcifications and Mass Lesions in Mammograms, IJCSNS *International Journal of Computer Science and Network Security*, 9(5), pp:255-265
6. Michael H. Torosian .**2002**. *Breast Cancer a Guide to Detection and Multidisciplinary Therapy,* Humana Press Inc. Totowa, New Jersey.pp33-37.
7. Feig S. and Yaffe MJ .**1995**. Digital Mammography, Computer-Aided Diagnosis and Telemammography, The Radiologic Clinics of North America, Breast Imaging, 33(6), PP:1205-1230.
8. Michael H. T. **2002**. "*Breast Cancer a Guide to Detection and Multidisciplinary Therapy*, Humana Press Inc. Totowa, New Jersey,p.60.
9. Kekre H.B., Tanuja K. Sarode and Saylee M. Gharge. **2009**. Tumor Detection in Mammography Images using Vector Quantization Technique, *International Journal of Intelligent Information Technology Application,* 2(5), pp:237-242.
10. Jelena B., Mario M., Kresimir D., and Mislav G. **2009** A Survey of Image Processing Algorithms in Digital Mammography, SCI 231, pp:631–657.
11. Al Mutaz .M. A., Safaai D., Nazar Z. . **2011**. Masses Detection in Digital Mammogram by Gray Level Reduction using Texture coding Method, *International Journal of Computer Applications*, 29(4), pp:19-23.
12. Haralick R.M., Shanmugam K., and Dinstein I., .**1973**. Textural Features for Image Classification, *IEEE Trans. on Systems, Man, and Cybernetics*, vol. Smc-3(6), pp:610-621.
13. Al Mutaz M. Abdalla, Safaai D. and Nazar Z. **2011**. Detection of Masses in Digital Mammogram Using Second Order Statistics and Artificial Neural Network, *International Journal of Computer Science & Information Technology* (IJCSIT), 3(3), pp:176-186
14. A Mohd. K., R Besar, WMD W. Z. and NN A. **2009**. Identification of masses in digital mammogram using gray level co-occurrence matrices, *Biomedical Imaging and Intervention Journal*, 5(3), pp:142-154.