Digital Imaging in Mammography towards Detection and Analysis of Human Breast Cancer

Prof. Samir Kumar Bandyopadhyay
Professor, Dept. of Computer Sc. & Engg,
University of Calcutta
92 A.P.C. Road, Kolkata – 700009, India

Indra Kanta Maitra[1]
Research Fellow, Dept. of Computer Sc. & Engg,
University of Calcutta
92 A.P.C. Road, Kolkata – 700009, India

ABSTRACT

Mammography is at present most popular and available method for early detection of breast cancer. The most common breast abnormalities that may indicate breast cancer are masses and calcifications. The challenge is to quickly and accurately overcome the development of breast cancer, which affects more and more women through the world. Masses appear in a mammogram as fine, granular clusters, which are often difficult to identify in a raw mammogram. Mammogram is one of the best technologies currently being used for diagnosing breast cancer. Breast cancer is diagnosed at advanced stages with the help of the mammogram image. In this paper, some simple segmentation processes have been develop to make a supporting tool to easy and less time consuming method of identification abnormal masses in mammography images. The identification technique is divided into four distinct parts i.e. preprocessing, selection, isolation and projection. The type of masses, orientation of masses, shape and distribution of masses, size of masses, position of masses, density of masses, symmetry between two pair etc are clearly sited after proposed method is executed on raw mammogram for easy and early detection of abnormality. The outcomes of the results are satisfactory and acceptable.

Keywords

Breast Cancer, Mammogram, Masses, GLCM, Contrast, Homogeneity, Energy.

1. 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 [3].

Mammogram is one of popular technique to identify 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 [7]. The small percentage of breast cancers that are not identified by mammography may be missed for just as mammography uses x-ray machines designed especially to image the breasts.

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. Segmentation is the fundamental process which partitions a data space into meaningful salient regions. Image segmentation essentially effects the overall performance of any automated image analysis system thus its quality is of the utmost importance.

Digital mammography is a technique for recording x-ray images in computer code instead of on x-ray film, as with conventional mammography. The first digital mammography [4] system received U.S. Food and Drug Administration (FDA) approval in 2000. An example of a digital mammography system is the Senographe 2000D. 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 [9].

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 [10].

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. Also, the appearance of an image may be compromised if there is powder or salve on the breasts or if you have undergone breast surgery.

Recent studies showed that the interpretation of the mammogram by the radiologists give high rates of false positive cases indeed the images provided by different patients have different dynamics of intensity and present a weak contrast. Moreover the size of the significant details can be very small. 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, 6, 11].

---

[1] Senior System Analyst, B.P. Poddar Institute of Engineering and Management, India.

Imaging techniques play an important role in helping perform breast biopsies, especially of abnormal areas that cannot be felt but can be seen on a conventional mammogram or with ultrasound [9]. In the paper some very simple processes have been develop to make a supporting tool to easy identification of abnormal masses in mammography images.

2. PROPOSED TECHNIQUE

This paper is basically concentrated to develop a tool to identify the abnormal growth of masses in breast using very simple algorithms. The tool will only identify the masses with some distinguishing features to ease further investigation. In this process we have selected the digital mammogram, which has become the most effective technique for early breast cancer detection. The mammogram images used in this experiment are taken from the mini mammography database of MIAS [12]. The original MIAS Database (digitized at 50 micron pixel edge) has been reduced to 200-micron pixel edge and clipped/padded so that every image is 1024 pixels x 1024 pixels. All images are held as 8-bit gray level scale images with 256 different gray levels (0-255) and physically in portable gray map (.pgm) format. The list is arranged in pairs of mammograms, where each pair represents the left and right breast of a single patient. In our experiment we have consider all types of breast tissues i.e. Fatty, Fatty-glandular, Dense-glandular and the abnormalities like calcification, well-defined or circumscribed masses, speculated masses and other ill-defined masses. We have considered more than hundred samples as a test case. The identification technique is divided into four simple parts i.e. preprocessing, selection, isolation and projection.

2.1 PREPROCESSING

Raw mammograms images are difficult to understand, thus a pre-processing phase is needed. The main objective of this process is to improve the quality of the image to make it ready for further processing by removing the irrelevant parts of the image. To reduce the time complexity of the process we manipulate the palette instead of the data part of the image. We derive mid index value of gray scale by calculating the max and minimum index value and eliminate the all index less than mid. We have continued it for twice and redistribute the index value uniformly. Through this color quantization method color difference turn into prominent and distinct.

Algorithm 1:
Input: Raw Mammogram Image (Image).
Isize = Size of the Image
Byte = Single byte of data of the image
MaxV = Maximum index value
MinV = Minimum index value
MidV = Mid value of MaxV and MinV
Phase = Count the process
MaxC = Maximum number of color in the phase
Output: Gray Scale Image with ¼ color shades.
Begin
Step1.   Read the Mammogram Image File.
Step2.   Loop I = 0, Isize
          Read Image Byte
          IF MaxV < Byte
                    MaxV = Byte

```
          End IF
          IF MinV > Byte
                    MinV = Byte
          End IF
          I = I + 1
          End Loop
Step3.    Loop Phase = 1, 2
          MidV = (MaxV + MinV) / 2
          MaxC = MaxV - MidV
          Loop I = MinV, MidV
                    Image. Palette [I]. Red = 0
                    Image. Palette [I]. Green = 0
                    Image. Palette [I]. Blue = 0
                    I = I + 1
          End Loop
          Loop I = MidV + 1, MaxV
                    Temp = Image. Palette [I]. Red - MaxC
                    Temp = temp * (MaxV / MidV)
                    Image. Palette [I]. Red = Temp
                    Image. Palette [I]. Green = Temp
                    Image. Palette [I]. Blue = Temp
                    I = I+1
          End Loop
          MinV = MidV
          Phase = Phase + 1
          End Loop
Step4. Write to Gray Scale Image
End
```

2.2 SELECTION

In this process, we have used Uniform color quantization method on palette, which is consisting of prominent and distinct color index into four-scale color to select the different region of mammogram prominently.

Algorithm 2:
Input: Gray Scale Image with ¼ color shades (Image1).
Ccode = Color index
Output: 4 Shade Gray Scale Image (Image2).
Begin
Step1.   Read the Gray Scale Image.
Step2.   Loop I = 0, 255
          Ccode = Image1. Palette [I]. Red
          IF Ccode <= 64
                    Ccode = 255;
          ELSE
                    IF Ccode <= 128
                              Ccode = 64
                    ELSE
                              IF Ccode <= 192
                                        Ccode = 128
                              ELSE
                                        Ccode = 192
                              End IF
                    End IF
          End IF
          Image2. Palette [I]. Red = Ccode
          IF Ccode < 255
                    Ccode = 0
          End IF

Image2. Palette [I]. Green = Ccode
Image2. Palette [I]. Blue = Ccode
    End Loop
Step3. Write to 4 Shade Gray Scale Image
End

2.3 ISOLATION

The selected portion is then isolated using homogeneous matrix method. In this process we have read image data 2X2 matrix pattern. Convert the index with red value of the color, which occur maximum in the matrix. Next we repeat the same by using 4X4 matrix. Now the whole mammogram image consist of some blocks with is homogeneous in nature. So, the different regions are become completely separated to each other and prominent due to use of color in the gray image. The edges of each region are clear.

Algorithm 3:
Input: 4-Shade Gray Scale Image.
Isize = Size of the Image
Output: Homogeneous Image
Begin
Step1.   Read the Gray Scale Image.
Step2.   Loop I = 1, Isize / 4
        Read 2X2 Data from 4-Shade Gray Image File to [Array 2 X 2]
        Check the maximum occurrence Index value.
        Copy maximum occurrence Index value to all cell of [Array 2 X 2]
        Write [Array 2 X 2] to Homogeneous Image
        I = I + 1
    End Loop
Step3.   Loop I = 1, Isize / 16
        Read 4 X 4 Data from Homogeneous Image to [Array 4 X 4]
        Check the maximum occurrence Index value.
        Copy maximum occurrence Index value to all cell of [Array 4 X 4]
        Write [Array 4 X 4] to Homogeneous Image
        I = I + 1
    End Loop
End

2.4 PROJECTION

The projection or fusion is simply to present the isolated portion by superimposing the color part of the image in original mammogram to measure the size, depth, density and magnitude of the masses. In addition, the symmetry of the two breast and architectural deformities can also be identified.

Algorithm 4:
Input: Raw Mammogram Image (Image1), Homogeneous Image (Image2).
Isize = Size of the Image
Output: Final Image
Begin
Step1.   Read the Raw Mammogram Image.
Step2.   Loop I = 64, 255
        Image1. Palette [I]. Red = I
        Image1. Palette [I]. Green = 0
        Image1. Palette [I]. Blue = 0
        I = I + 64
    End Loop
Step3.   Write to Final Image.
Step4.   x = 1, y = 1
Step5.   Loop I = 0, I <= Isize
        Byte = Read the Raw Mammogram Image.
        R = Image. Palette [Byte]. Red;
        IF R MOD 64 = 0
            Byte = Byte + 1
        End IF
        Byte = Read the Homogeneous Image.
        R = Image2.Palette [Byte]. Red
        IF R MOD 64 = 0 AND R != 0
            Byte = R
        End IF
        IF x MOD 16 = 0 OR y MOD 16 = 0
            Byte = 255
        End IF
        IF x MOD 1024 = 0
            y = y + 1
        End IF
        x = x + 1
        Write to Final Image.
        I = I + 1
    End Loop
End

3. TEST RESULT

The success of the proposed technique is determined by the extent to which potential abnormalities can be extracted from analogous mammograms based on analysis of their image. The Mammographic Image Analysis Society Database is used to evaluate the proposed technique. More than hundred bilateral image pairs were used for testing. A randomly selected set bilateral pairs drawn from the database pairs with calcification, circumscribed masses, speculated masses and other ill-defined masses speculated and circumscribed lesions was used for the same to obtain result.

Major objective of the algorithms is to remove the non-masses area from the mammogram to identify the only the masses or presence of abnormality clearly. The stage, intensity, type, future and treatment can only be detected on the basis of type of masses, orientation of masses, shape and distribution of masses, size of masses, position of masses, density of masses, symmetry between two pair etc. The outputs of aforesaid algorithms are depicted in the following (see figures 1 to 10) mammograms of normal breast and breast with masses along with the histogram and colormap of the mammogram images.
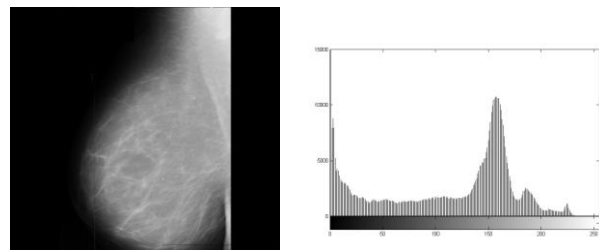

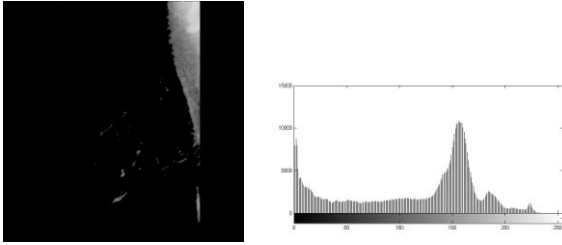Figure 1.   Normal Breast Mammogram with Histogram and Colormap.

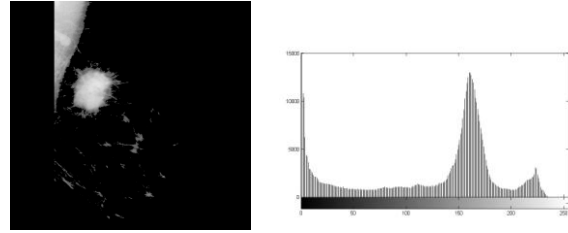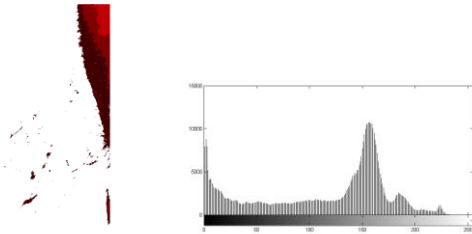Figure 2.   Normal Breast Mammogram after Preprocessing with Histogram and Colormap.

Figure 3.   Normal Breast Mammogram after Selection process with Histogram and Colormap.

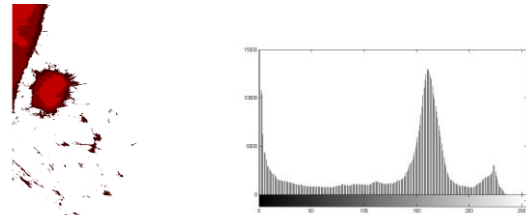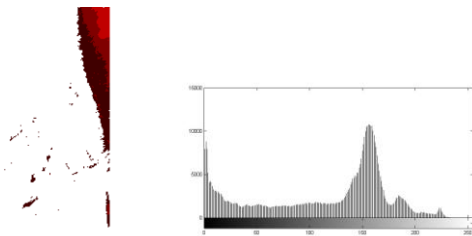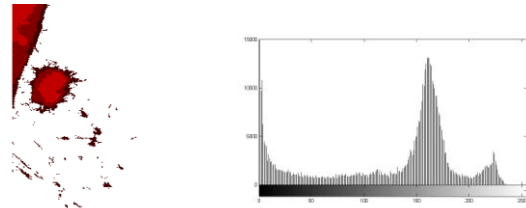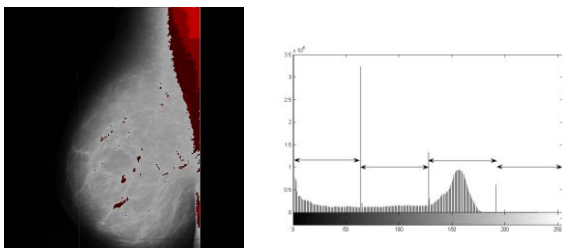Figure 4.   Normal Breast Mammogram after Isolation process with Histogram and Colormap.

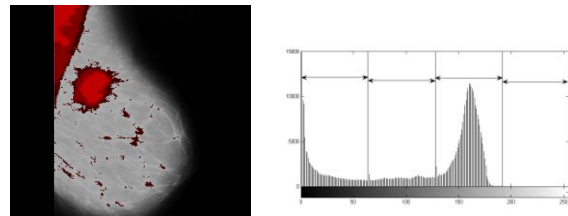Figure 5.   Normal Breast Mammogram after Projection process with Histogram and Colormap.

Figure 6.   Breast Mammogram with masses with Histogram and Colormap.

Figure 7.   Breast Mammogram with masses after Preprocessing with Histogram and Colormap

Figure 8.   Breast Mammogram with masses after Selection process with Histogram and Colormap.

Figure 9.   Breast Mammogram with masses after Isolation process with Histogram and Colormap.

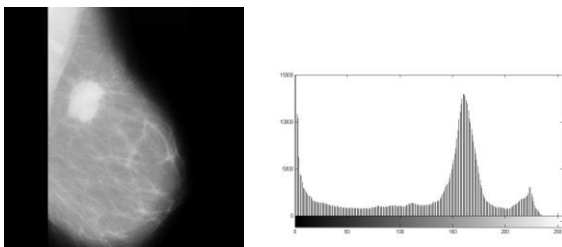Figure 10. Breast Mammogram with masses after Projection process with Histogram and Colormap.

In the following images (see figure 11) enlarge view of masses is stated. From the same type of masses, orientation of masses, shape and distribution of masses, size of masses, position of masses, and density of masses is clearly visible.
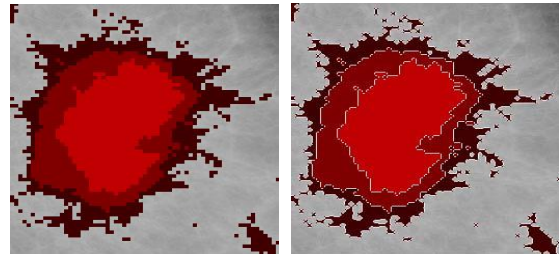

Figure 11. Enlarge view of Mammogram after Projection Process showing Abnormal Masses.

In the following (see figure 12) Mammogram after Projection process showing Abnormal Masses in X-axis and Y-axis measuring area, shape and size.
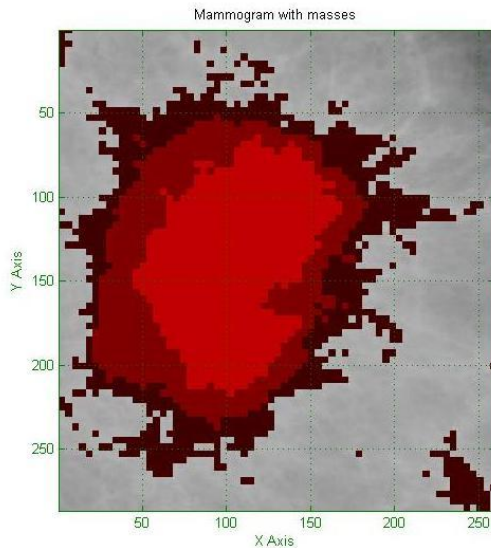

Figure 12. Enlarge view of Mammogram after Projection Process showing Abnormal Masses in X-axis and Y-axis measuring area, shape and size.

In the next image (see figure 13) pair after the Projection process showing the comparative view of symmetry.
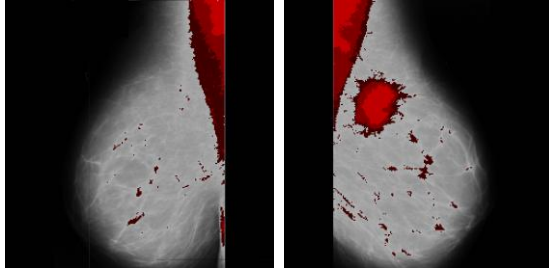

Figure 13. Mammogram after projection of two pair showing comparative view.

In the subsequent image pair (see figure 14) after the Projection process showing the comparative view of size, position, density and distribution of abnormal masses.
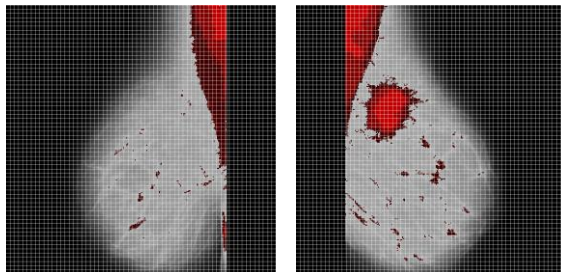

Figure 14. Mammogram after projection of two pair showing comparative view of size, position, density and distribution masses.

In the edge map of image pair (see figure 15) after the Projection process showing the comparative view of symmetry.
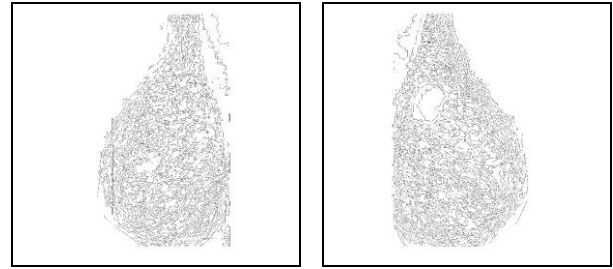

Figure 15. Edge comparison of pair after Projection.

4. RESULT ANALYSIS

Texture features have been proven to be useful in differentiating masses and normal breast tissues [1, 2, 13]. Texture features are able to isolate normal and abnormal lesion with masses and micro-calcifications. In the experimental work, the texture features are extracted using gray level co-occurrence matrices (GLCM). The matrices are constructed at a distance of d = 1 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 [1, 2, 8, 13].

The texture descriptor derived from GLCM is contrast, energy, homogeneity and correlation of gray level values. The contrast measures the amount of local variations present in an image, while energy is the sum of squared elements in GLCM. Energy may also be referred as uniformity or the angular second moment. The homogeneity descriptor refers to the closeness of the distribution of elements in GLCM to the GLCM diagonal.

Based on the database and resultant images derived from algorithm are tested with the range of values of contrast, homogeneity and energy of masses and non-masses tissues of 8x8 tile area are shown in Table I, Table II and Table III, respectively. The graphical presentations of the Contrast, Homogeneity and Energy values are also sited in Figure 16, Figure 17 and Figure 18, respectively. It is observed that the values of contrast, homogeneity and energy for image containing masses and image containing non-masses are highly discriminated. This has proven the usefulness of the proposed method using three texture descriptors in differentiating the masses and non-mass mammogram.

Table 1. Table Contrast Value of Masses and Non Masses.

	Contrast at direction θ			
	0º	45º	90º	135º
Mass	0.0806	0.1215	0.0685	0.1153
Non-Mass	0.0667	0.0865	0.0459	0.0916

Table 2. Homogeneity Value of Masses and Non Masses.

	Homogeneity at direction θ			
	0º	45º	90º	135º
Mass	0.9808	0.9728	0.9811	0.9756
Non Mass	0.9829	0.9797	0.9852	0.9772

Table 3. Energy Value of Masses and Non Masses.

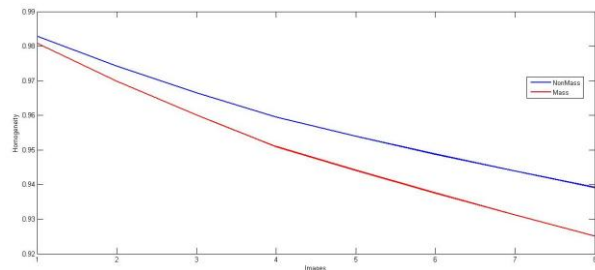|	Energy at direction θ			
	0º	45º	90º	135º
Mass	0.3455	0.3421	0.3463	0.3435
Non Mass	0.3462	0.3447	0.3478	0.3436
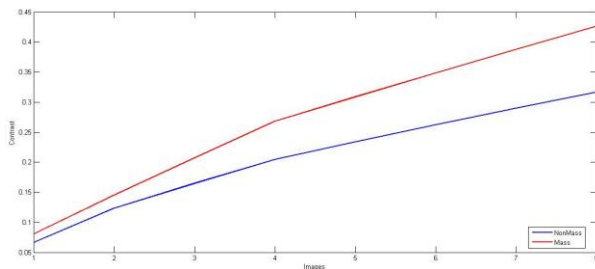
Figure 16. Contrast value at θ =0º

Figure 17. Homogeneity value at θ =0º
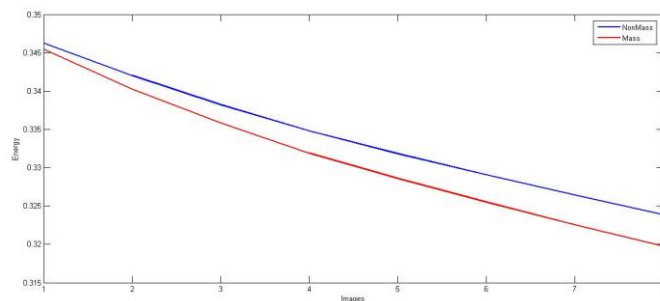

Figure 18. Energy value at θ =0º

5. CONCLUSION

It can be very difficult to decide who may have a breast cancer and who may have a non-cancerous breast condition. Advances in computing and telecommunications have resulted in the availability of a range of tools for use in mammography quality assurance and support system. The majority focuses on either enabling mammography to examine and diagnose cases, or providing image archives that serve as reference material. Limited emphasis has been placed on analysing the diagnostic process used by mammography to reach a diagnosis and using this as a resource for improving diagnostic performance. This method has potential for further development because of its simplicity that will motivate online or real-time breast cancer diagnosis in providing the opinion.

6. REFERENCES

[1] Ball JE. Digital mammogram spiculated mass detection and spicule segmentation using level sets. Proceedings of the 29th Annual International Conference of the IEEE EMBS. 2007: 4979-84.

[2] Bovis K, Singh S. Detection of masses in mammograms using texture features. 15th International Conference on Pattern Recognition. 2000: 267-70.

[3] Breast Cancer Facts & Figures, 2009-2010, American Cancer Society, Inc.

[4] FDA Web site, http://www.accessdata.fda.gov/scripts/ cdrh/cfdocs/ cfMQSA/mqsa.cfm [last visited on September 30, 2009].

[5] R. Highnam and M. Brady, Mammographic Image Analysis, Kluwer Academic Publishers, 1999. ISBN: 0-7923- 5620-9.

[6] Dr. H. B. Kekre, Tanuja K. Sarode and Saylee M. Gharge, "Tumor Detection in Mammography Images using Vector Quantization Technique", International Journal of Intelligent Information Technology Application, 2009, 2(5):237-242.

[7] Michaelson J, Satija S, Moore R, et al. The pattern of breast cancer screening utilization and its consequences. Cancer. Jan 1 2002; 94(1): 37-43.

[8] A Mohd. Khuzi1, R Besar, WMD Wan Zaki, NN Ahmad1,"Identification of masses in digital mammogram using gray level co-occurrence matrices", Biomedical Imaging and Intervention Journal, http://www.biij.org/2009/3/e17.

[9] National Cancer Institute (NCI) Web site, http://www.cancernet.gov [last visited on September 30, 2009].

[10] RadiologyInfo.org developed jointly by Radiological Society of North America and American College of Radiology.

[11] E. E. Sterns, "Relation between clinical and mammographic diagnosis of breast problems and the cancer/ biopsy rate," Can. J. Surg., vol. 39, n°. 2, pp. 128-132, 1996.

[12] J Suckling et al (1994): The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica. International Congress Series 1069 pp375-378.

[13] Yang SC, Lin YJ, Hsu GC et al. Mass screening and features reserved compression in a computer-aided system for mammograms. IEEE World Congress on Computational Intelligence. 2008: 3779-86.