



PRIMARY RESEARCH

## Automatic detection of breast cancer in mammogram images

Faozia Ali S. Alsarori <sup>1\*</sup>, Reza Hassanpour <sup>2</sup><sup>1</sup> Yıldırım Beyazıt University, Ankara, Turkey<sup>2</sup> Çankaya University, Ankara, Turkey

### Index Terms

Automatic Breast Cancer Detection  
Computer Aided Diagnosis  
Mammogram Images  
Artificial Neural Networks

**Received:** 21 September 2016**Accepted:** 22 October 2016**Published:** 19 December 2016

**Abstract**— The aim of this study is to harness the great potential of image processing techniques which have evolved significantly in the last years, to build an automatic system to detect and diagnose breast cancer in the digital mammographic images in order to help those interested people in this domain, such as radiologists and specialists in oncology and to improve their performance by reducing error rates of breast cancer diagnosis. As long as segmentation and extracting the effective features of mammograms play a major role to isolate and classify suspicious regions which can be subject to cancer, in this work, we focus on abnormality detection using multi-thresholding OTSU's method to segment the Region Of Interest (ROI). Then the texture features of the segmented ROI are extracted which are used to classify the ROI as normal or abnormal tissue by using an Artificial Neural Network (ANN). This system can correctly classify the tested region by a rate of 93.80%.

© 2017 The Author(s). Published by TAF Publishing.

### I. INTRODUCTION

Among the various kinds of cancer, Breast Cancer (BC) is one of the most dangerous cancers [1], [2] and [3] also it is considered the second leading cause of death among women aged between 45 and 55 years, for early detection, mammography imaging is one of the most effective tools because of its ability to detect abnormalities more than 2 years before they are obvious. In fact, a tumor at the early stage appears as tiny shiny spots in the mammographic images [4]. These spots are calcium sediments. Its scientific label is Microcalcifications (MCs) [5] and [6]. MCs have a subtle nature, in most cases unclear in the images and difficult to identify them where this constitutes a big challenge for radiologists due to the nature of human vision system. Conversely, the second type of breast cancer called Masses may be easier to detect due to the nature of the size, shape and color contrast and at the same time some types of masses can be difficult to detect as a result of tissue nature, which may look similar to normal breast tis-

sue (parenchyma). In order to ameliorate the radiologists' performance which has been leading to increase the treatment efficiency and the survival chance at the same time, reducing the misdiagnosis, sought a lot of specialists to other technologies to help the radiologists in reading mammographic images and identify suspicious regions in the breast and provide a second opinion for them to detect microcalcifications and masses earlier and to diagnose if those MCs or masses are benign or malignant. These technologies are Computer Aided Systems which supply better detection and diagnosis techniques (Computer Aided Detection (CADe) and Computer Aided Diagnosis (CADx)). This study proposes an automatic system to detect and diagnose abnormalities in digital mammography images.

### II. LITERATURE SURVEY

[7] have applied region segmentation after using morphological operations, a set of features has been computed from each of the segmented areas. In this work 57

\* Corresponding author: Faizia Ali S. Alsarori

† Email: alsarorifa@gmail.com

mammogram images were tested containing breast masses from MIAS data set, including speculated, circumscribed, and ill-defined masses. The result of using such method has reached 80% sensitivity rate at 2.3 as false-positives per image.

[8] proposed extended Sobel, Prewitt and Kirsch edge operators to locate and segment tumor from the mammography images. This method can detect the location and edges of the tumor clearly better than using Sobel, Prewitt and Kirsch edge operators, as well as using proposed extended 5x5 Sobel operator with mammography images giving better segmentation than extended Kirsch and Prewitt operator.

[9] applied a study represented in collecting mammography image to 58 biopsies to indicate breast cancer, in which each mammogram was taken 10 -18 months earlier than breast cancer detection to evaluate them. Marking two regions, first one corresponding to a malignant mass area

developed later, the second indicates the same region in the first region but on the normal breast, then 62 features of texture and photometry have been calculated. This method can detect breast cancer by nearly 72%.

[10] presented an adaptive genetic programming for masses' classification. The performance of the genetic programming classifier was good in discriminating between malignant and benign masses with registered 99.5% accuracy in training and 98% in testing.

### III. APPLIED METHODS AND PROPOSED SYSTEM

This work mainly introduces a system to detect abnormality in digital mammographic images using image enhancement techniques for eliminating noise as well as to obtain a better image contrast in order to segment the suspicious region and figure out its features to train a classifier. Figure 1 shows the steps of the proposed system.

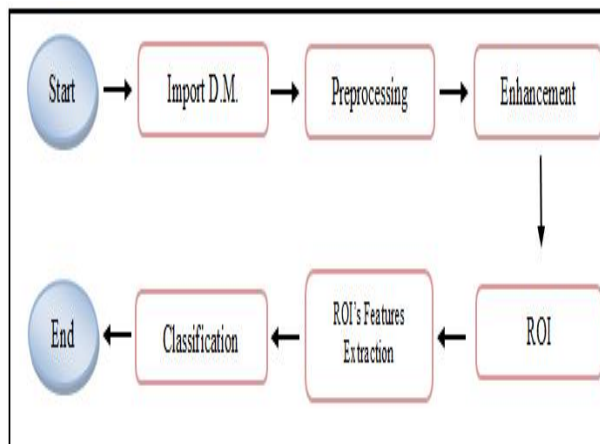


Fig. 1 . Flow chart digram of the proposed system

#### A. Preprocessing

The goal of preprocessing is decreasing image objects as well as enhancing the breast's edges before extracting (ROI). Each stage of preprocessing is explained below:  
Enhancement

##### -Unsharp masking

$$g1(x, y) = g(x, y) + f1(x, y)$$

$$f1(x, y) = 4g(x, y) - g(x - 1, y) - g(x + 1, y) - g(x, y - 1) - g(x, y + 1)$$

##### -Histogram modification:

##### Exponential modification approach

$$g2(x, y) = \begin{cases} 3x + 5y + z \\ 7x - 2y + 4z \\ - 6x + 3y + 2z \end{cases}$$

#### B. Segmentation

##### Multi-thresholding OTSU's method:

This method is used for dividing the image into several classes. Thus, the work of this technique depends on the nature of the lesion contrast. So, applying this method to segment the abnormality requires careful choosing of the number of the classes.

### C. Feature Extraction

Features are often used to point the numerical information group that is pertinent to solve the computational tasks related to certain applications. Moreover, they play a very effective role to detect the mammographic image abnormalities due to its nature.

### D. Feature Selection

In this step, probability density function (PDF) has been used to select the best features among the 15 features that were already extracted. The selected features are:

1. Optical Density Sum: the measurement of the integrated abnormality optical density is un-normalized value.
2. Optical density variance: the variance function value of the abnormality optical density is normalized value.

#### 3. Standard Deviation:

The deviation describes the dispersion within a local region. It is determined using the following:

$$\sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (p(i, j) - \mu)^2}$$

Where  $\sigma$  estimates the square deviation of the grey pixel value  $p(i, j)$  mean.

#### 4. Smoothness:

$$R = 1 - 1/(1 + \sigma^2)$$

Where  $R$  the measurement the gray level contrast that can be utilized to create smoothness descriptors.

#### 5. Skewness:

A number only describes the form of the distribution while  $S$  describes the level of distribution of a pixel asymmetry in the particular window around the pixel mean.

$$S = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left( \frac{p(i, j) - \mu}{\sigma} \right)^3$$

Where  $p$  is the pixel gray value at point  $(i, j)$ ,  $\sigma$  is the standard deviation and  $m$  is the mean.

#### 6. Kurtosis:

$K$  calculates the flatness and peakness of a distribution proportional to the distribution in normal case.

$$K = \left\{ \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left( \frac{p(i, j) - \mu}{\sigma} \right)^4 \right\} - 3$$

Where,  $p(i, j)$  is the pixel value at point  $(i, j)$ ,  $m$  and  $\sigma$  are the Mean and Standard Deviation respectively. The -3 term

makes the value zero for a normal distribution.

#### Sum Variance:

This feature puts relatively high weights on the elements that differ from the average value of  $P(i, j)$ .

$$SV = \frac{1}{MN} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu)^2 p(i, j)$$

### E. Classification

A neural network was designed to train the classifier where the candidate regions which have already been decided as abnormal or normal were utilized.

## V. EXPERIMENTS AND RESULTS

### A. Experiments

In this experiment we test our methods on mammographic image containing a malignant mass.

#### Step I : Preprocessing:

##### 1. Noise Removal

Here media filter was applied on the mammographic image (mdb271) for removing the light problem which is shown as dark lines in Figure 2.

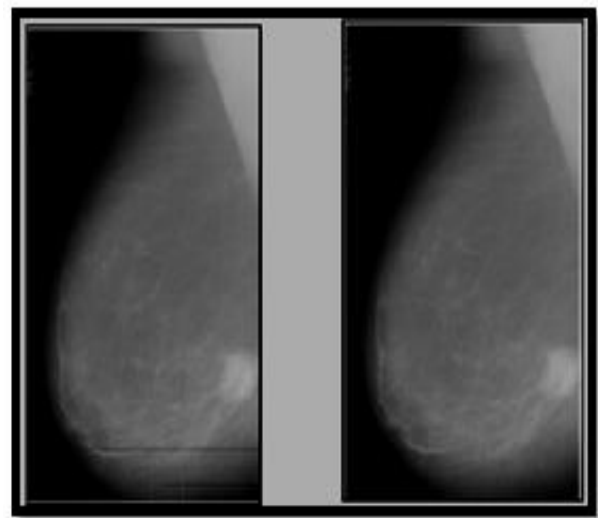


Fig. 2. The result of using the 2D media filter

Extra image's objects are removed in this step by using morphological thresholding and contrast enhancement techniques. Figure 3 shows the sub-steps results in this process.

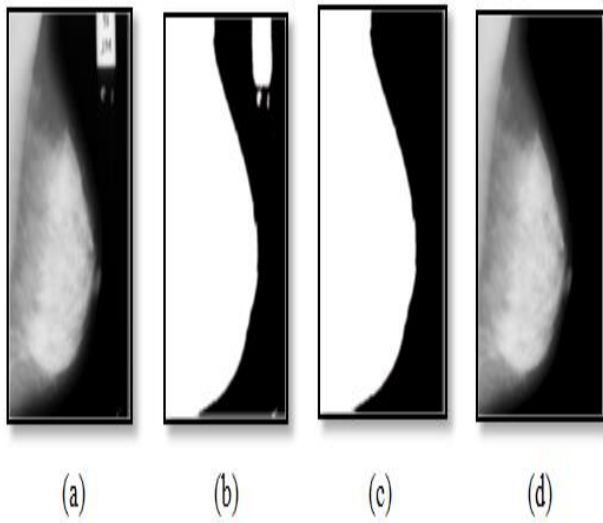


Fig. 3 . (a) Original Image, (b) The result after using global threshold, (c) The result after keeping the object with the largest number of pixels and (d) The final Image

3. Pectoral Muscle Removal

In this step pectoral muscle is removed by using contrast enhancement and geometrical techniques.

• **Step II : Enhancement:**

This step prepares the image for ROI segmentation by applying image enhancement techniques. Figure 4 shows the resulting image and the mass (suspicious region) appears lighter.

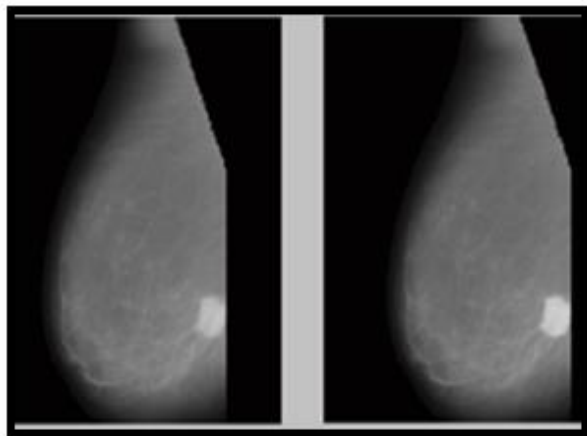


Fig. 4 . The result of applying enhancement techniques

• **Step III : Segmentation:**

In this stage of system’s work, ROI (abnormal region) is extracted by using multi-thresholding OTSU. Figure 5 shows the result of multi-thresholding OTSU.

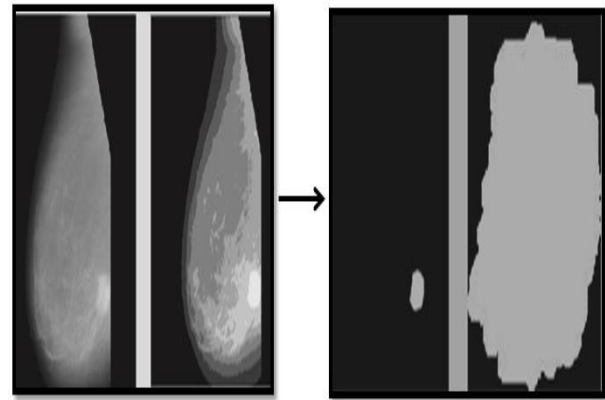


Fig. 5 . Applying multi thresholding OTSU’s method, the segmented region is shown

• **Step IV : Feature Extraction:**

After getting the features’ values as shown in Figure 6 for all the mammogram images, a new value (attribute) has been added to the crated dataset referred to as normality or abnormality.

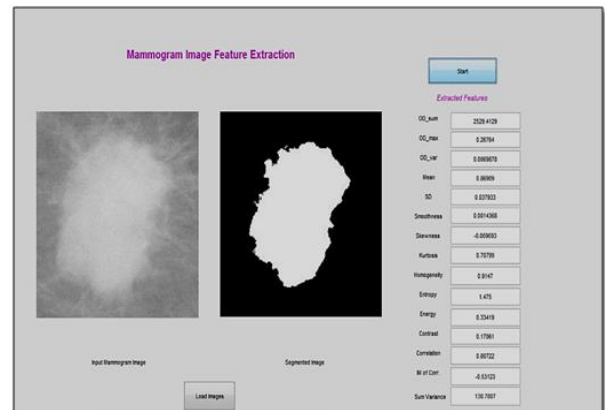


Fig. 6 . Snapshot of output result of the feature extraction

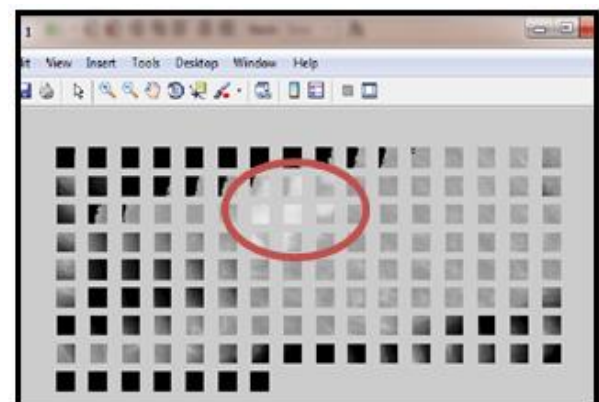


Fig. 7 . Snapshot of sub-images (mdb184)

|   |          |        |         |          |           |        |          |          |          |       |
|---|----------|--------|---------|----------|-----------|--------|----------|----------|----------|-------|
|   | 1        | 2      | 3       | 4        | 5         | 6      | 7        | 8        | 9        | 10    |
| 1 | NaN      | NaN    | NaN     | NaN      | NaN       | NaN    | NaN      | NaN      | NaN      | NaN   |
| 2 | NaN      | NaN    | 0.00085 | 0.01312  | 0.0378    | 0.976  | 1        | NaN      | NaN      | NaN   |
| 3 | NaN      | NaN    | 0.999   | 0.999073 | 0.944     | 0.002  | 6.11E-05 | 0.001    | 7.29E-05 | 0.034 |
| 4 | 0.0430   | 0.435  | 0.370   | NaN      | NaN       | NaN    | 5.92E-05 | 0.845    | 0.9814   | 0.999 |
| 5 | 0.999    | 0.001  | 0.0001  | 7.36E-05 | 0.0023    | 0.001  | 0.003266 | 0.0747   | 0.99     | 0.415 |
| 6 | 0.759    | 0.442  | 0.231   | 0.120    | 0.9998006 | 0.9998 | 0.157377 | 4.32E-05 | 0.0005   | 0.003 |
| 7 | 7.75E-05 | 0.0001 | 0.0048  | 0.013    | 0.3287565 | NaN    | NaN      | 0.213    | 0.483    | 0.999 |

Fig. 8 . Part of the results after classifying a divided image

**B. Classification**

Classification is performed by using trained classifier for each part where one tested image is divided into sub images (16 × 16 parts). Figure 7 shows cancerous regions surrounded by the red circles and Figure 8 shows part of the result after classifying all sub images to appropriate class either around 0 or 1.

**C. Evaluation**

After getting the statistical values shown in Table 1 (a) and (b), we could evaluate the

TABLE 1  
AFTER GETTING THE STATISTICAL VALUES SHOWN IN TABLE 1 (A) AND (B), WE COULD EVALUATE THE

| No.     | Total of sub images | Real abnormality | TP | FP | FN | TN  | P  | N   |
|---------|---------------------|------------------|----|----|----|-----|----|-----|
| mdb 271 | 149                 | 9                | 9  | 12 | 0  | 128 | 9  | 140 |
| mdb 184 | 135                 | 10               | 8  | 12 | 2  | 115 | 10 | 127 |
| mdb 165 | 66                  | 14               | 14 | 5  | 0  | 47  | 14 | 52  |
| mdb 150 | 152                 | 17               | 17 | 3  | 0  | 132 | 17 | 135 |
| mdb 134 | 201                 | 5                | 5  | 6  | 0  | 190 | 5  | 196 |
| mdb 132 | 159                 | 17               | 15 | 2  | 2  | 142 | 17 | 144 |
| mdb 111 | 106                 | 22               | 22 | 2  | 0  | 82  | 22 | 84  |

Could Evaluate the proposed system by getting the following values: Average of CM = Sensitivity = 98.3134%, Average of CR = 74.997%, Average of Specificity = 92.28%

and Average of Accuracy = 93.8059%  
Table 1(a)  
Statistical values of the system performance

TABLE 2  
STATISTICAL VALUES OF THE SYSTEM PERFORMANCE

| NO.     | CM=TPR   | FPR      | CR=PPV   | F1       | ACC      |
|---------|----------|----------|----------|----------|----------|
| mdb 271 | 1        | 0.085714 | 0.428571 | 0.6      | 0.919463 |
| mdb 184 | 0.8      | 0.094488 | 0.4      | 0.533333 | 0.89781  |
| mdb 165 | 1        | 0.096154 | 0.736842 | 0.848485 | 0.924242 |
| mdb 150 | 1        | 0.022222 | 0.85     | 0.918919 | 0.980263 |
| mdb 134 | 1        | 0.030612 | 0.454545 | 0.625    | 0.970149 |
| mdb 132 | 0.882353 | 0.013889 | 0.882353 | 0.882353 | 0.975155 |
| mdb 111 | 1        | 0.02381  | 0.916667 | 0.956522 | 0.981132 |

**V. CONCLUSION**

Due to the enormous number of the breast anomalies' properties and the nature of the human visual system, it is normal that some abnormalities can be missed or misclar-

ified. This work focuses on extracting the suspicious region with finding useful features, multi-thresholding algorithms for detecting and extracting the abnormality then collection of the photometric and texture features that were extracted

to build the classifier. The proposed system has achieved a sensitivity of 98.3134%, a correctness of 74.997%, a specificity of 92.28% and an accuracy of 93.8059%.

#### REFERENCES

- [1] M. Sameti, R. K. Ward, J. Morgan-Parkes and B. Palcic, "Image feature extraction in the last screening mammograms prior to detection of breast cancer," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, 46-52, 2009.  
**DOI: 10.1109/JSTSP.2008.2011163**
- [2] S. Niha, B. Jantarasiriput, N. Tonyongdalaw and N. Vaichompu, "Reproductive health among bangoebadae Muslim women: Cervical cancer care," *International Journal of Health and Medical Sciences*, vol. 2, no. 3, pp. 52-57, 2016.  
**DOI: 10.20469/ijhms.2.30002-3**
- [3] E. I. Putri, R. Magdalena, L. Novamizanti, "The detection of cervical cancer disease using an adaptive thresholding method through digital image processing," *Journal of Advances in Health and Medical Sciences*, vol. 1, no. 1, pp. 30-36, 2015. **DOI: 10.1016/j.compbiomed.2005.12.006**
- [4] F. A. S. Alsarori, "Automatic detection of breast cancer in mammogram images," Doctoral dissertation, 2013.
- [5] C. Varela, P. G. Tahoces, A. J. Méndez, M. Souto and J. J. Vidal, "Computerized detection of breast masses in digitized mammograms," *Computers in Biology and Medicine*, vol. 37, no. 2, pp. 214-226, 2007.
- [6] H. M. Kim and K. A. Bae, "Electronic and optical properties of Indium Zinc oxide thin film prepared by using Nanopowder target," *Japanese Journal of Applied Physics*, vol. 50, no. 4R pp. 5103-5108, 2011.  
**DOI: 10.7567/JJAP.50.045801**
- [7] A. R. Dominguez and A. K. Nandi, "Detection of masses in mammograms via statistically based enhancement, multilevel thresholding segmentation, and region selection," *Computerized Medical Imaging and Graphics*, vol. 32, no. 4, pp. 304-315, 2008.
- [8] H. B. Kekre and S. M. Gharge, "Image segmentation using extended edge operator for mammographic images," *International journal on computer science and Engineering*, vol. 2, no. 4, pp. 1086-1091, 2010.
- [9] M. Sameti, R. K. Ward, J. Morgan-Parkes and B. Palcic, "A method for detection of malignant masses in digitized mammograms using a fuzzy segmentation algorithm," in *Engineering in Medicine and Biology Society, Proceedings of the 19th Annual International Conference of the IEEE*, vol. 2, 1997, pp. 513-516.
- [10] R. J. Nandi, A. K. Nandi, R. Rangayyan and D. Scutt, "Genetic programming and feature selection for classification of breast masses in mammograms," in *28th Annual International Conference of the IEEE*, 2006, pp. 3021-3024.

— This article does not have any appendix. —