Efficient Breast Cancer Classification Using Tumor Area and Boundaries Calculation

Belal A. Elmonem[1], Taha M. Mohamed[2], and Sherif Kholeif[3],
[1]Belal_1111@hotmail.com, [2]tahamahdy3000@yahoo.com, [3]sherifkholeif@yahoo.com
[1,2,3] Faculty of Computers and Information, Helwan University, Egypt.

Abstract— **Many of the women across the world suffer from breast cancer. Since the causes of this disease remain unknown, early detection and diagnosis are essential to control the breast cancer, increase the success of the treatment, save live, and reduce costs. Unfortunately, the methods proposed in the literature are suffering from low detection rates. In this paper, we proposed two new methods called C-Comp and CRoI. Both methods efficiently detect benign and malignant breast cancer tumors using shape circularity. The experimental results show that, the proposed C-Comp method achieved a detection rate of 12.3% higher than other previous work. However, the proposed CRoI method achieved a detection rate of 5.6 % higher than other previous work.**

Keywords: Breast cancer, Tumor Circularity, Tumor Classification, and Tumor Detection.

I. INTRODUCTION

THE Breast cancer is one of the deadliest diseases in human history. In 2012, it was reported that, about 40,000 of deaths, due to breast cancer, occur in the United States. Cancer starts when one cell changes and starts growing. Then, the cell is divided rapidly out of control forming a tumor. Middle-aged women are more prone to have breast cancer. In Arab world, statistics show that, 40,000 persons die each year due to the cancer [1].

The symptoms of breast cancer include breast lumps, accompanied by changes in size and skin texture or color. Breast cancer spreads quickly if it is not being controlled. Early detection and diagnosis is essential to control breast cancer, which increases the success of treatment, saves lives, and reduces costs. Researchers make great efforts to find solutions to this problem [2].

Mammography is one of the diagnostic tools, used by doctors, to detect the tumors [2]. In some cases, the available mammograms are of low quality due to mammography limitation factors. In this case, when using low quality images, the classification of the tumor is very difficult leading to classification ambiguity.

In this case, the tumor is almost indistinguishable. Usually, the tumor takes a regular shape in the beginning of the disease.

Then, the tumor is evolving to an irregular shape [3]. This is illustrated in Fig. 1 [2].
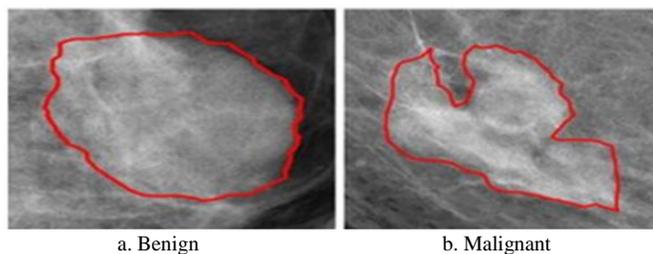

a. Benign   b. Malignant
Fig 1: Tumor Types [2]

In this paper, we propose two new methods for classifying the breast tumor to either benign or malignant. The proposed methods are called C-Comp and CRoI. The two methods try to detect the shape regularity in order to enhance the automatic classification.

This paper is organized as follows; section I is the introduction. Section II presents the necessary background. Section III presents the literature review. Section IV introduces the proposed methods. Section V shows the experimental results. Finally, the paper is concluded in section VI.

II. BACKGROUND

Breast tumors have many forms. Such forms may include circular, oval, regular, or even irregular shape [3] as shown in Fig.2. Tumors are classified to two types; benign or malignant. Often, the benign tumor cells are one of regular shapes, and it is a mass with limited and surrounding soft boundaries. However, the malignant tumors are often of irregular shapes, and it possesses irregular boundaries.
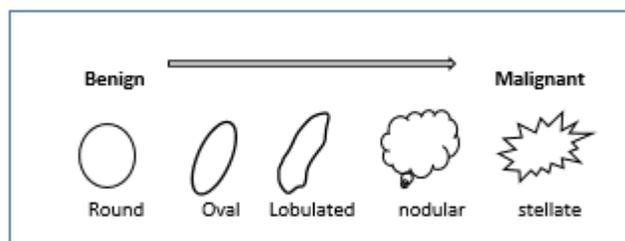
Fig.2: Morphological Shapes of Mammographic Images [3].

Most of the research in breast cancer classification is oriented to detect the regularity of the cell boundaries in order to correctly classify the cell to either benign or malignant. This section reviews some of the necessary background and concepts used to compute and detect the tumor type and its classification based on its shape.

A. The Correlation Coefficient

The correlation coefficient is used to compute the relationship between the two coordinates of each point on the boundary of the tumor. Equation 1 shows the formula used to calculate the correlation coefficient [4].

$$Corr = \frac{cov\ xy}{\sigma x * \sigma y} \tag{1}$$

Where cov xy is covariance matrix, σx is the standard deviation of the x dimension, and σy is the standard deviation of the y dimension. The Covariance matrix is computed using Equation 2 [4].

$$Cov = \frac{\sum xy}{n} \tag{2}$$

Where n is the length of each vector. Usually, the value of the correlation coefficient can determine the tumor type as shown next in the results section.

B. The Cmax method

In [5], the authors propose the Cmax method. In this method, a circle is drawn outside the borders of the shape in order to localize the ROI. Consequently, both the area of the shape and the area of circle are calculated.

The result of dividing shape area by circle area can determine the tumor type whether it is benign or malignant.

If the value is greater than 60%, then it is being classified as benign; otherwise it is being classified as a malignant [5]. Equation 3 shows the formula used to compute the Cmax method.
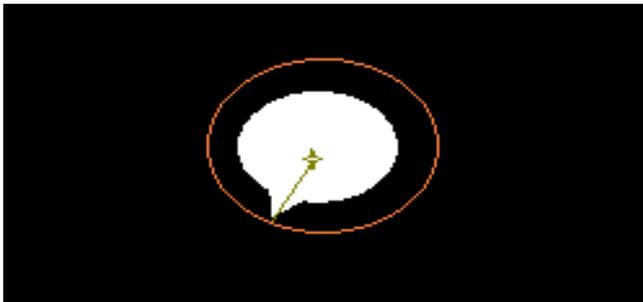

Fig. 3: The Circularity of the Tumor [5].

$$C_{max} = \frac{A_{shape}}{A_{Circle}} * 100 \tag{3}$$

Where A_{shape} represents the area of the shape, A_{Circle} represents the area of the circle as shown in Fig. 3.

III. LITERATURE REVIEW

In [4], the correlation coefficient is used to classify the tumor type. In [5], the Cmax algorithm is proposed to classify the tumor images to either benign or malignant. In [2], the Contour method is proposed by converting the tumor image from 2D to 1D signature. In [6], the region of interest is extracted by using the spherical wavelet transform. In [7], the image is firstly segmented and enhanced before classification. In [8], statistical parameters like entropy, mean, energy, correlation, texture, and standard deviation are used. Then, these parameters are fed into the neural network. In [9], three types of data mining are proposed in order to classify the data, and to predict the breast cancer type. In [10], some Thresholding methods are evaluated, such as Otsu, Valley Emphasis Method, and neighborhood valley emphasis. In [11], the contour algorithm is used, and the advanced K-means algorithm is also used. In [12], the segmentation with threshold is used to extract the ROI. In [13], a calculation of the fractal dimension are used. In [14], the convexity estimator is used to analyze the shape. However, the paper assumes only a polygon shape of the tumor, which is not always true. In [15], a multi-resolution analysis is used to analyze the image. However, this technique is limited when dealing with noisy images.

IV. THE PROPOSED METHODS

We proposed two new methods called C-Comp and CRoI. The target of these two methods is the early detection of the breast cancer by detecting the regularity of the tumor boundaries.

A. C-Comp

The C-Comp used the concept of compactness to identify the shape of the tumor, and hence classify the tumor. The proposed C-Comp method consists of six stages as shown in Fig. 4.
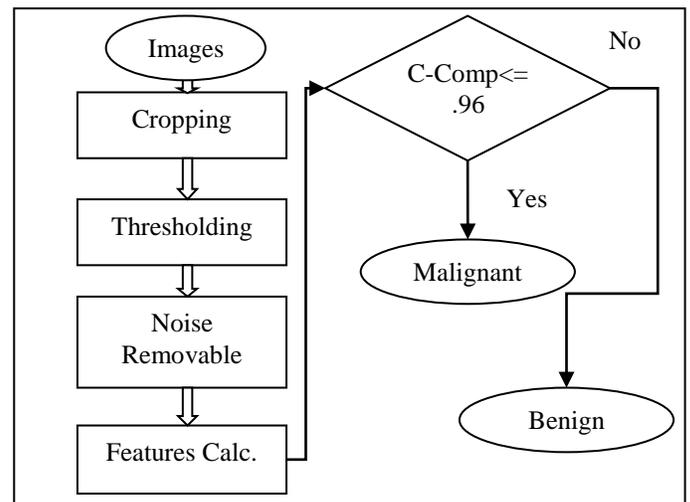

Fig.4: Flowchart for the C-Comp Algorithm

First, the tumor is cropped from the mammogram image to ease the classification process.

The cropping process is very important to localize the region of interest (ROI), and hence, extract the different features used in classification. Fig. 5 shows two different cases having different types of tumor [16]. The determination of tumor locations is specified according to the MAIS dataset determination [16].
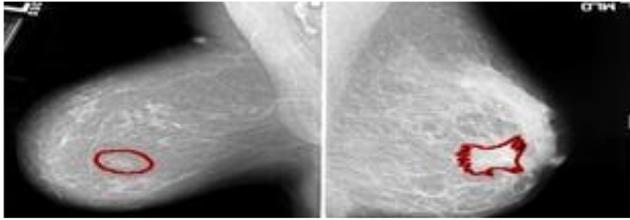
a. Benign   b. Malignant
Fig.5: Example from MIAS dataset [16].

Next, a Thresholding and noise removal is performed. Usually, this noise is a type of salt and pepper noise. Often, the median filter is used to remove this type of noise. This stage is very important because it helps to distinguish the ROI. In this stage, a threshold T is used to convert the gray scale image into a binary one. The resulted image contains only two values black and white. The result of Thresholding is shown in Fig. 6.
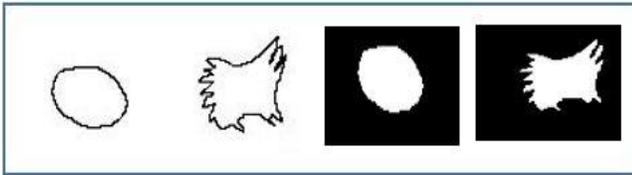

Fig.6: Thresholding and Noise Removal

After previous preprocessing operations, the compactness formula is applied. It is a relationship between tumor area and its perimeter. The value of the compactness is then compared to a value of 0.96 which is the value that differentiates between tumor types. This choice of this value is determined based on our experiments. If the value of the compactness is less than or equal to 0.96, then the tumor is classified as malignant. Otherwise, it is classified as benign. The compactness formula is shown by equation 4 [17].

$$Com = \frac{4\pi . A_{area}}{P_{(perimeter)^2}} * 100 \tag{4}$$

Where is the shape area. ( ) is the shape perimeter. In Fig. 7, the pseudocode of the algorithm is illustrated.

Pseudocode:	Algorithm C-Comp
Input:	Images
Output:	Tumor Classification

1- Read MIAS Image Ii
2- Use Manual Image Cropping Ri
3- Make Thresholding (Ri>T)
4- Romove any noise
5- Apply feature extraction
6- If Result>.96 then
     Tumor is Benign
     Else the tumor is malignant

Fig.7: Pseudo code of the C-Comp algorithm

B. CRoI method

It consists of the stages shown in Fig. 8. The algorithm begins by cropping the input image. Then, the image is thresholded, and converted to binary image.
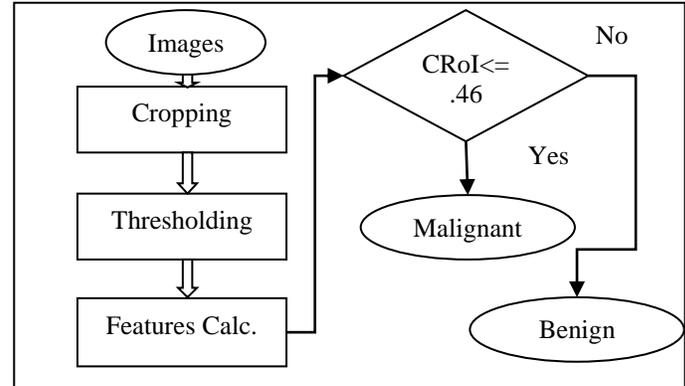

Fig.8: Flowchart of the CRoI Algorithm

Then, we calculate the center of gravity of the tumor, which is called the centroid. After computing the centroid, we draw the circle surrounding the ROI. Then, the circum and area are also calculated. Next, we use Equation 5 which represents the proposed CRoI. Equation 5 measures CRoI by subtracting the area of the ROI from the area of the circle, then the result is divided by the area of the circle.

$$CRoI = \frac{C_a - R_a}{C_a} * 100 \tag{5}$$

Where C_a represents the area of circle and R_a represents the area of the ROI. If the CRoI ratio is greater than or equal to 46%, then the tumor is classified as malignant. Otherwise, the tumor is classified as benign. This threshold is also obtained through our experimentation. Fig. 9 shows the algorithm to illustrate the steps.

Pseudocode:	Algorithm CRoI
Input:	Images
Output:	Tumor Classification

1- Read MIAS Image Ii
2- Use Manual Image Cropping Ri
3- Apply feature extraction
4- Calculate the centroid of ROI
5- Calculate the circle area
6- Calculate the ratio
7- If Result>.96 then Tumor is Benign
     Else the tumor is malignant

Fig. 9: Pseudocode of the CRoI algorithm

V. THE EXPERIMENTAL RESULTS

The experiments are conducted on 1000 images from MIAS Dataset [16]. We use MATLAB ver.15, core I5 processor, and 8 GB ram. We used 500 benign and 500 malignant images. The images are tested in samples. Each sample contains 200

image (100 of them are malignant, and the other 100 are benign). The accuracy values are computed using Equation 6.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} * 100 \qquad (6)$$

Where TP and TN are true positive and true negative. FP and FN are false positive and false negative. Four methods are experimented. Two methods are our proposed C-Comp and CRoI, and the other two are previous work. All experiments are compared and tabulated. Results show that, the proposed C-Comp method achieves a higher accuracy than the previous work described in [4].

However, its accuracy is less than the Cmax method proposed in [5]. In order to achieve higher results, we proposed the new CRoI method. In this case, the CRoI method achieves the highest results among all tested methods as shown in Table1.

Table1: Comparison of the Experimental Results

S.	C-Comp Proposed			Previous work Saed Hamoud[4]		
	Ben. TP	Mal. TN	ACC.%	Ben. TP	Mal. TN	ACC. %
S1	73	75	74	73	59	66
S2	80	60	70	63	54	58.5
S3	66	71	68.5	71	46	58.5
S4	71	58	64.5	69	44	56.5
S5	72	69	70.5	47	46	46.5
Avg.			69.5			57.2
S.	CRoI Proposed			Previous work Cmax[5]		
	Ben. TP	Mal. TN	ACC.%	Ben. TP	Mal. TN	ACC. %
S1	79	68	73.5	50	91	70.5
S2	89	76	82.5	73	86	79.5
S3	80	78	79	45	94	69.5
S4	75	70	72.5	43	90	66.5
S5	76	74	75	48	93	70.5
Avg.			76.5			71.3

Table 1 contains five samples from S1 to S5. The choice of each sample images is random. However, each sample contains 100 benign images and 100 malignant images. Four methods are compared; C-Comp, previous work method [4], CRoI and Cmax. Each method contains three columns. Columns are the number of benign (TP), the number of malignant (TN) and the accuracy (Acc). The averages of all results are computed in the last row.

After comparing all results, it is noted that, that C-Comp method exceeds the accuracy of the previous work in [4] in detecting the type of the tumor. On the average, the C-Comp method achieved 69.5% in detection accuracy. The correlation method achieves 57.2% in accuracy, with a difference of 12.3% between them in favor of the C-Comp method. Shown in Fig. 10 the accuracy bar chart between the C-Comp and the
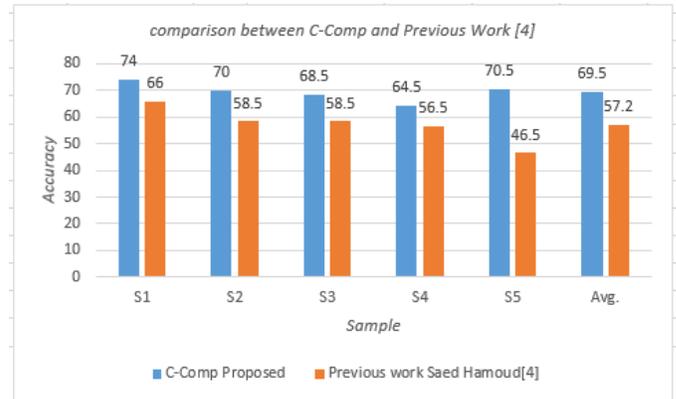
previous work described in [4].

Fig.10 Comparison between C-Comp and previous work [4]

It is noted from Fig. 10 that, the C-Comp exceeds the correlation coefficient in accuracy with respect to all samples. However, the C-Comp method was less than the Cmax method [5]. So, we proposed the new CRoI method. The CRoI method is compared to the Cmax proposed in [5]. Moreover, the CRoI method exceeded the accuracy for Cmax [5] in detecting the type of the tumor. On the average, the CRoI achieved 76.5% in accuracy. The Cmax achieves 71.3% in accuracy, with a difference of 5.2% between them in favor of CRoI. In Fig. 11, the bar chart shows the comparison between the CRoI and the Cmax methods from accuracy point of view.
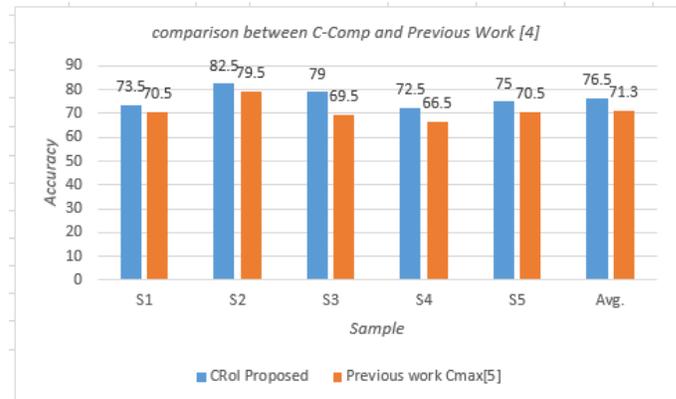

Fig.11 Comparison between Proposed CRoI and Cmax [5]

It is noted from Fig. 11 that, the CRoI method exceeds the accuracy of the Cmax method [5]. On the average, the CRoI achieves 76.5% in detection accuracy, whereas the Cmax method achieves only 71.3 % in detection accuracy.

VI. CONCLUSION

This paper presents two new methods for breast cancer classification called C-Comp and CROI. The experimental results show that, the C-Comp and the CRoI methods achieves accuracy rates higher than other previous work. Moreover, the C-Comp outperforms other previous work by 12.3%. Additionally, the CRoI method outperforms other previous work by 5.2%. The CRoI method achieves a rate of 7% higher than the C-Comp. Therefore, the CRoI method achieves the

highest accuracy rates among the studied methods.

REFERENCES

[1]   C. Freitas, "Automatic analysis of mammography images: enhancement and segmentation techniques", *Engineering Faculty, Engineering Faculty, Proto University,* Master Thesis, p. 152, 2011.

[2]   H. Li, X. Meng, T. Wang, Y. Tang, and Y. Yin, "Breast masses in mammography classification with local contour features", Biomedical Engineering journal, vol. 16, issue 1, 2017.

[3]   H. Eittah, H. Awed, and Z. Bukhary, "Effect of Health Education on Raising Female Students", Journal of Natural Sciences Research, vol. 4, 2014.

[4]   S.Hamouda, Reda H Abo El-Ezz, and Mohammed E Wahed, "Enhancement Accuracy of Breast Tumor Diagnosis in Digital Mammograms", Journal of Biomedical Sciences, vol. 6, issue 4, 2017.

[5]   T.K.Heo, "Breast Cancer Classification of Mammographic Masses Using Circularity Max Metric, A New Method", Master Thesis, South Dakota State University, 2016.

[6]   P.Görgel, "Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines", The Journal of Knowledge Engineering, vol. 32, issue 1, 2015.

[7]   Angayarkanni N , Kumar D, and Arunachalam G, "The Application of Image Processing Techniques," journal of pharmaceutical science and research, vol. 8, no. 10, 2016.

[8]   S. Naranje, "Early Detection of Breast Cancer using ANN", International Journal of Computer Sciences and Engineering, vol. 4, issue 7, 2016.

[9]   V. Chaurasia and S. Pal, "A Novel Approach for Breast Cancer Detection Using Data Mining Techniques", international Journal of Innovative Research in Computer and Communication Engineering , vol. 2, no. 1, 2014

[10] M. Al-Bayati and A. El-Zaart, "Mammogram Images Thresholding for Breast Cancer", Scientific Research, 2013.

[11] C. Saipriya,M.Therasa, "Automatic Tumour Detection In Mammogram Using Supervised", International Journal of Advance Research, vol. 3, no. 2, 2017.

[12] J. Anitha, J. Dinesh Peter, "Mammogram segmentation using maximal cell strength updation in cellular automata", Medical & Biological Engineering & Computing, vol. 53, no. 8, 2015.

[13] Alan Chan, J. Tuszynsk, "Automatic prediction of tumor malignancy in breast cancer with fractal dimension", Royal Society, vol. 3, no. 12, 2016.

[14] J. Zunic, P.L. Rosin, "A new convexity measure for polygons", vol. 26, no. 7, 2004.

[15] E. A. Rashed, "Multiresolution mammogram analysis in multilevel decomposition", Pattern Recognition letters, vol. 28, issue 2, 2007.

[16] "Digital Database for Screening Mammography", University of Florida, Available at, last visited 6-April-2019 http://www.eng.usf.edu/cvprg/Mammography/Database.html,

[17] M. Wirth, "Shape Analysis & Measurement," University of Guelph, Canada, Ph.D. thesis, 2004.