

Using Shape Circularity for Breast Cancer Detection

Abed S. Alsahli^{1*}, Belal A. Elmonem², Taha M. Mohamed^{2,3} and Sherif Kholeif^{2,3}

¹Faculty of Business, University of Jeddah, KSA, Saudi Arabia; asalshle@uj.edu.sa

²Faculty of Computers and Information, Helwan University, Egypt, Saudi Arabia;

Belal_1111@hotmail.com, tahamahdy3000@yahoo.com

³Faculty of Business, University of Jeddah, KSA, Saudi Arabia, sherifkholeif@yahoo.com

Abstract

Background/Objectives: This study aims to automate breast cancer detection to either benign or malignant. This process helps physicians for early detection of the tumor. **Methods Analysis:** We propose a novel method for breast cancer detection called CARA. The proposed method is based on Circular ROI (Region of interest) of the tumor. The circularity of the tumor determines the tumor type either benign or malignant. **Findings:** The proposed CARA method achieves high accuracy rates compared to other previous work. CARA exceeds other previous work by more than 20% in accuracy. Moreover, CARA outperforms other previous work in all performed trials. Additionally, the proposed CARA method outperforms other previous work with respect to execution time. **Improvements/Applications:** Automatic classification of breast cancer to either benign or malignant in order to help physicians for early detection of the tumor.

Keywords: Breast Cancer Detection, Circularity, Tumor Area and Tumor Shape

1. Introduction

The Breast cancer is the cancer that forms in breast cells. It is the most common type of cancer among women. Breast cancer may affect both men and women, but it is more common in women. Breast cancer is diagnosed in 12% of women in the United States over the course of their lifetimes, and more than 250,000 new cases of breast cancer were diagnosed in the United States in 2017¹. Additionally, half a million women have developed cancer. Also, 100,000 women die every year in Europe due to breast cancer². In Arab world, breast cancer is dangerous. Records associated with breast cancer cases diagnosed from 2001 to 2008 among Saudi women indicated that, 6,922 females with breast cancer were registered in the Saudi Cancer Registry³.

The highest prevalence was found in eastern region of KSA, followed by³. Several factors may be the causes of breast cancer. These factors differ for different geographical locations³. Breast cancer is lower among men than among women, however, men were at more risk of death than women. Moreover, divorced patients were more at risk of

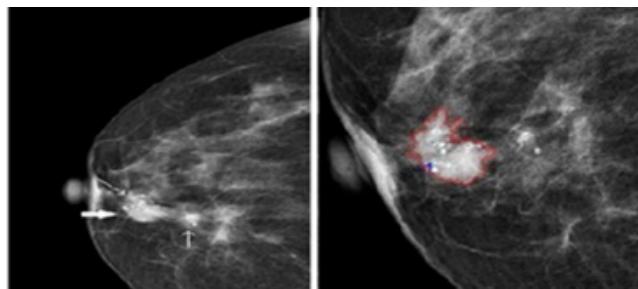
death than married, single, and widowed patients, though there was high mortality among married patients³. Unfortunately, Saudi Arabian women with breast cancer who are divorced may have worse consequences and less support than other women because of the conservative nature of the country⁴.

Breast cancer starts in different parts of the breast. Most breast cancers begin in the ducts that carry milk to the nipple. Some other cancers start in the glands that make breast milk⁵. The knowledge of breast cancer risk factors can strongly contribute to breast cancer prevention efforts³.

The treatment of breast cancer is very effective especially when being initiated at an early age and sustained over a lifetime³. There are two types of breast calcifications, Micro-Calcification Clusters (MCCs) and mass lesions (Macro-calcification). Micro-calcification is a small of calcium that has amassed in the breast. It appears in the X-Ray as a small bright plug as shown in Figure 1. Often, it appears as points on the mammogram, which are considered as benign tumors. Often, in Macro-calcification, breast cells are transformed into abnormal

*Author for correspondence

(malignant) and spread to other areas of the patient body. The difference between the two types is shown in Figure 1⁶.

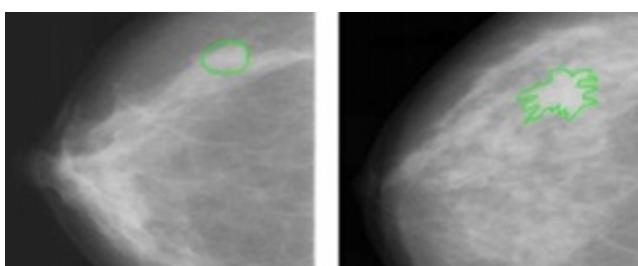


A. Micro-calcification B. Macro-calcification

Figure 1. Difference between the two types of the tumor⁶.

In most cases, benign tumors are one of semi-circle or a circle shaped, and it is often a mass with soft limit. However, malignant tumors often have complicated shapes, and have irregular boundaries as shown in Figure 2⁷.

Mammography is one of the tools used by doctors for breast cancer diagnosis. Moreover, mammography was used for early detection of breast cancer. Physicians found difficulty in diagnosing a mammography. This may be due to lack of clarity of images. Therefore, the computer was used to analyze mammogram images for more accuracy⁸. Image processing, data mining, and machine learning techniques are often used to classify the mammograms to either benign or malignant in order to achieve early detect of the tumor. The proposed method aim to help physicians to achieve early detecting of the disease, and to find the suitable treatment as well. In this study, we propose a new method for breast cancer detection. The proposed method is called CARA which is a modification of the method called Circularity Max Metric (Cmax)⁹. The proposed method used dataset of images from the MIAS dataset¹⁰.



A- Benign

B-Malignant

Figure 2. The difference between Benign and Malignant tumors⁷.

This study is organized as follows; Section 1 presents the introduction, Section 2 presents the necessary background. Section 3 surveys the literature review. In Section 4, the proposed method is explained. Section 5 presents the experimental results. Finally, the study is concluded in Section 6.

2. Background

Breast cancer may come in many forms. Some of them are circular, oval, or regular shaped. However, the irregular shaped is also common. These shapes are shown in Figure 3⁸.

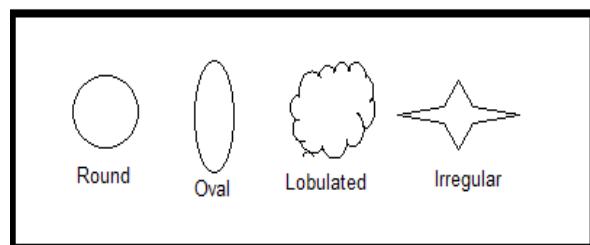


Figure 3. The difference between benign and malignant tumors¹¹.

Much of the research in breast cancer depends on the regularity of the boundaries of the tumor in order to correctly classify the tumor. For example, regular tumor shapes are often indicating to benign tumors. However, irregular shapes are often indicating to malignant tumors. Here, we will review a previous work method that called Circularity Max Metric (CMAX)⁹. CMAX uses the shape of the tumor to classify the tumor to either benign or malignant. Our proposed method is a modification of the CMAX method. The Cmax⁹ method is used for classifying the tumor to either benign or malignant. In⁹, the authors enhance the mammogram images as a preprocessing step. Then, they used the Wiener filter to remove the noise from the images. So, some features are extracted from the images such as the centroid of the shape. Next, a surrounding circle is drawn to the shape as shown in Figure 4. The surrounding shape is used for computing the area of the circle and the shape. Equation (1) is used in this computation⁹.

$$\text{Circularity} = \frac{A_{\text{Shape}}}{A_{\text{Circle}}} \quad (1)$$

From Equation (1), represents the area of the shape. The represents the area of the circle. The result of division of both areas determines the tumor type either benign or malignant. If the result is greater than 65%, then the tumor is classified as benign, otherwise it is malignant.

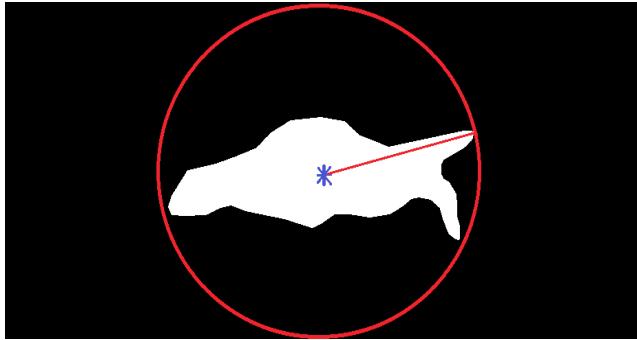


Figure 4. Illustrate Circularity Max (Cmax)⁹.

3. Literature Survey

In¹¹, some features are extracted from image. Then, the correlation coefficient is computed on the Region of Interest (ROI). In⁹, the authors present the CMAX method. In¹², a comparison between four thresholding methods is conducted to detect breast cancer. In¹³, breast cancer is detected by using both segmentation and thresholding. This method is based on a focus of the pixels on image. In both researches of^{3,4}, the authors describe and analyze some risk factors related to breast cancer disease in Saudi Arabia. Their conclusion is that, most of the worldwide risk factors are still applicable in Saudi Arabia. However, the Saudi case has more risk factors such as over-weight and the high percentage of divorce. In¹⁴, the authors applied morphological filter to enhance image. In¹⁵, the Spherical Wavelet Transform (SWT) is used, after segmentation, and extracting of the ROI. However, these techniques are limited when dealing with noisy images. In¹⁶, the calculation of the fractal dimension was used. In¹⁷, the authors present two new methods for breast cancer detection based on tumor shape.

4. The Proposed Method

Here, we will present the proposed method which is called CARA. The goal of the proposed method is the early detection of breast cancer. The proposed CARA method is compared with Circularity Max Metric (CMAX) method⁹. The proposed CARA algorithm modifies the

previous Cmax⁹ method. The flowchart of the proposed CARA method consists of six stages as shown in Figure 5.

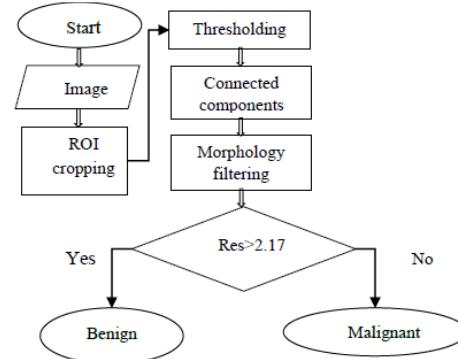


Figure 5. Flowchart of the CARA algorithm.

The CARA algorithm begins as follows; first, an image from the data set is selected. The second stage is the specification of the Region of Interest (ROI). The third stage is the Thresholding. The fourth stage is the using of the connected components algorithm¹⁸. To find the unlabeled pixel, the flood-fill algorithm¹⁹ is applied. The fifth stage is the noise removal by using the morphological filtering²⁰. The sixth stage is the features extraction process in order to classify the tumor. The feature extraction phase requires measuring the shape area and the circle area which surround the shape, as shown in Figure 6.

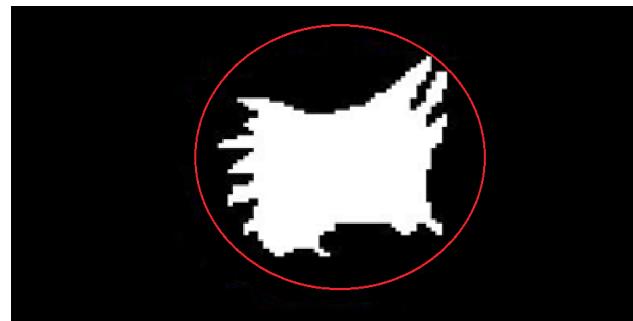


Figure 6. Localizing the tumor by using a circle.

Here the features extracted are the shape area and the circle area. After calculating the area of the shape and the area of the circle, Equation (2) is used to compute the result Res.

$$Res = \frac{Circle_A}{Circle_A - ROI_A} \quad (2)$$

Where represents the area of circle and represents the area of the ROI. If Res >= 2.17, then the tumor is classified as benign, otherwise the tumor is classified as malignant.

The steps of the CARA algorithm is detailed in the pseudocode shown in Figure 7.

Pseudocode: Algorithme CARA
Input: Image
Output: Tumor classification
1- Select Image Ii
2- Get ROI Ci
3- features extraction
4- Calculate the centroid of RoI
5- Calculate the circle of area
6- Calculate the ROI of area
7- Calculate the res
8- If res > = 2.17 then tumor is benign
9- Else the tumor is malignant

Figure 7. Pseudocode of the CARA algorithm.

5. Experimental Results

In our experimental results, 1200 images are tested. A PC with MATLAB ver.17, core I5 processor, and 8 GB ram is used in the experimentation. The used images are categorized as 600 of them are benign and the other 600 are malignant. The test images are divided into six samples. Each sample contains 200 images (100 benign, 100 malignant). The accuracy is calculated by using Equation (3).

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (3)$$

Where TP is the true positive. The true positive refers to the number of correctly classified malignant cases. TN is the true negative. The true negative refers to the number of correctly classified benign cases. FP and FN are false positive and false negative respectively. They are referring to the misclassified cases in either malignant or benign test cases.

5.1 Comparison of CARA and the Previous Work in¹¹

Table 1 shows a comparison of the experimental results between the proposed method and the previous work method proposed in¹¹. In the table, the first column refers to the sample number. In the second column Ben. TN Refers to the number of true negative cases (benign) is specified. In the third column, the number malignant true positive cases is specified. The Acc. refers to the accuracy

computed from Equation (3). Finally, the Avg. column refers to the average.

From Table 1, it is noted that, the proposed CARA method exceeds the previous work method proposed in¹¹ in accuracy. On the average, the CARA achieved accuracy about 76.9%. However, the method proposed in⁸ achieved average accuracy of only 56.8%, with a difference of accuracy between them about 20%.

Table 1. Comparison between CARA proposed and previous work in¹¹

Sample	CARA (Proposed)			Previous work [11]		
	Ben. TN	Mal. TP	Acc.%	Ben. TN	Mal. TP	Acc.%
1	78	69	73.5	72	58	65
2	88	77	82.5	64	52	58
3	81	77	79	70	47	58.5
4	75	71	73	67	45	56
5	77	75	76	50	47	48.5
6	80.8	74.2	77.5	60	50	55
Avg.			76.9			56.8

5.2 Comparison of CARA and the Previous Work Method Cmax⁹

The accuracy of the proposed CARA method is also compared with accuracy of the previous Cmax method presented in⁹. Table 2 shows the result of the comparison between them. Again, it is noted from Table 2 that, the proposed CARA method exceeds the previous work method in accuracy by about 5.6%. Moreover, the proposed method achieves higher accuracy rates in all trials. Therefore, the CARA method achieves the highest accuracy rates among all studied methods.

Table 2. Comparison between CARA method and previous work Cmax⁹

Sample	CARA Proposed			previous work Cmax ⁹		
	Ben. TN	Mal. TP	Acc.%	Ben TN	Mal. TP	Acc.%
1	78	69	73.5	51	90	70.5
2	88	77	82.5	74	85	79.5
3	81	77	79	46	93	69.5
4	75	71	73	42	91	66.5
5	77	75	76	49	92	70.5
6	80.8	74.2	77.5	52	91	71.5
Avg.			76.9			71.3

From the figure it is noted that, while the Cmax method⁹ outperforms the method presented in¹¹, however, the proposed CARA method outperforms both method presented in^{9,11}. Thanks to the CARA method.

Figure 8 shows a graphical comparison between the CARA method and the previous work in both^{9,11}.

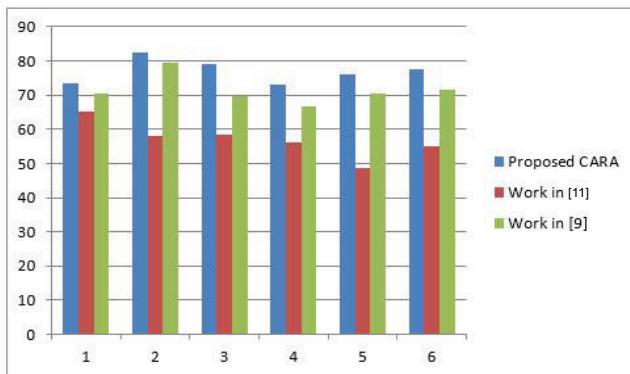


Figure 8. Comparison of accuracy between Proposed CARA both previous work^{9,11}.

5.3 Comparison of CARA and the previous work method CROI¹⁷

Although the accuracy of the algorithm presented in our previous work in¹⁷ is almost the same as CARA, however, the computation time of the proposed CARA algorithm is interesting. The proposed algorithm reduces the computation time compared to our previous method, CROI, presented in¹⁷. The equation used for CROI computation¹⁷ is shown by Equation (4) as:

$$CRoI = \frac{C_a - R_a}{C_a} \times 100 \quad (4)$$

When comparing both Equations (2) and (4), it is noted that, Equation (4) used multiplication of a factor 100 for obtaining percentage. In fact, this multiplication does not affect the accuracy, it is just a scaling factor. However, the multiplication process is a time consuming process. So, we neglect this scaling term in the computation of the proposed Equation (2) of CARA. This is proved using the conducted experimental results shown in Table 3. In Table 3, five images are used for experimentation. The execution time, of both algorithms, is computed. The time difference is also calculated and tabulated in Table 3.

Table 3. Comparison of algorithm time computation of both CROI and CARA

Tested Image	CRoI Time	CARA Time	Time Difference
image 1	0.000061	0.00006	1E-06
image 2	0.000062	0.00006	0.000002
image 3	0.000062	0.000059	3E-06
image 4	0.000062	0.00006	0.000002
image 5	0.000062	0.000061	0.000001

From Table 3, it is noted that, in all cases, the proposed CARA method has lower execution time compared to CROI. The time difference is also illustrated in Figure 9.

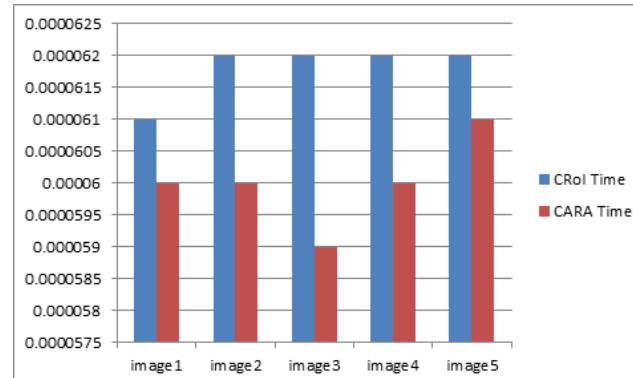


Figure 9. The execution Time difference between CARA and previous work¹⁷.

6. Conclusion

This study presents an automated technique for breast cancer for early detection. The proposed method, is called CARA, achieves an accuracy rate of 76.9% on the average. When compared to two methods from the literature, the proposed method outperforms other previous work by more than 20% on the average. Moreover, in all trials, the proposed method outperforms other previous methods. Additionally, the proposed CARA method outperforms previous work with respect to the execution time of the algorithm.

7. References

- Yasmin M, Waks AG, Winer EP. Breast cancer treatment: A review. The Journal of the American Medical Association (JAMA). 2019; 321(3):288-300. <https://doi.org/10.1001/jama.2018.19323> PMid:30667505

2. Breast cancer management for surgeons. Available from: <https://www.springer.com/gp/book/9783319566719>
3. Breast cancer in Saudi Arabia and its possible risk factors and control. Available from: https://www.researchgate.net/publication/331595710_Breast_Cancer_in_Saudi_Arabia_and_its_Possible_Risk_Factors_and_Control_Chapter_4_Recent_Studies_Advances_in_Breast_Cancer
4. Breast cancer mortality in Saudi Arabia: Modeling observed and unobserved factors. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0206148>
5. About breast cancer. Available from: <https://www.cancer.org/content/dam/CRC/PDF/Public/8577.00.pdf>
6. Breast cancer: diagnostic imaging and therapeutic guidance. Available from: <https://www.amazon.com/Breast-Cancer-Diagnostic-Therapeutic-Guidance/dp/3132019313>
7. Tang J, Rangayyan RM, Xu S, El Naqa I. Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances. *IEEE Journal*. 2014; 13:2-17.
8. Balasubramanian S. Mammogram mass classification using various geometric shape and margin features for early detection of breast cancer. *International Journal of Medical Engineering and Informatics*. 2012; 4(1):1-20. <https://doi.org/10.1504/IJMEI.2012.045302>
9. Breast cancer classification of mammographic masses using circularity max metric- A new method. <https://pdfs.semanticscholar.org/c5ba/14bc4c419134e51d9052d59b8957873b5865.pdf>
10. Digital database for screening mammography. Available from: https://www.researchgate.net/publication/228543259_The_Digital_Database_for_Screening_Mammography.
11. Hamouda S, Abo El-Ezz RH, Wahed ME. Enhancement accuracy of breast tumor diagnosis in digital mammograms. *Journal of Biomedical Sciences*. 2017; 6(4):1-8. <https://doi.org/10.4172/2254-609X.100072>
12. Al-Bayati M, El-Zaart A. Mammogram images thresholding for breast cancer detection using different thresholding methods. *Advances in Breast Cancer Research*. 2013; 2: 72-7. <https://doi.org/10.4236/abcr.2013.23013>
13. Anitha J, Dinesh Peter J. Mammogram segmentation using maximal cell strength updating in cellular automata. *International Federation for Medical and Biological Engineering*. 2015; 53(8):1-8. <https://doi.org/10.1007/s11517-015-1280-0> PMid:25841356
14. Morphological operations for the mammogram image to increase the contrast for the efficient detection of breast cancer. Available from: https://www.researchgate.net/publication/265948354_Morphological_Operations_for_the_Mammogram_Image_to_Increase_the_Contrast_for_the_Efficient_Detection_of_Breast_Cancer
15. Görgel P. Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines. *The Journal of Knowledge Engineering*. 2015; 32(1):155-64. <https://doi.org/10.1111/exsy.12073>
16. Automatic prediction of tumor malignancy in breast cancer with fractal dimension. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5210682/>
17. Efficient breast cancer classification using tumor area and boundaries calculation. Available from: <http://fcihib.fci.helwan.edu.eg/index.php/Informatics-Bulletin/login?source=%2Findex.php%2FInformatics-Bulletin%2Fissue%2Fview%2F2>
18. Klusowski JM. Estimating the number of connected components in a graph via subgraph sampling. *Pattern Recognition*. 2018; 1-44.
19. Tjiharjadi S, Wijaya MC, Setiawan E. Optimization maze robot using a * and flood fill algorithm. *International Journal of Mechanical Engineering and Robotics Research*. 2017; 6(5):366-72. <https://doi.org/10.18178/ijmerr.6.5.366-372>
20. Li H, Meng X, Wang T, Tang Y, Yin Y. Breast masses in mammography classification with local contour features. *Biomedical Engineering Journal*. 2017; 16(1):44. <https://doi.org/10.1186/s12938-017-0332-0> PMid:28410616 PMCid: PMC5391548