**ORIGINAL PAPER**

# Automatic suspicions lesions segmentation based on variable-size windows in mammography images

Bahram Sadeghi[1] · Meysam Karimi[1] · Samaneh Mazaheri[2]

**ABSTRACT**

Breast cancer is the second main cause of death in women of western countries, so early detection and prevention is crucial. Early detection increases the likelihood of treatment as well as patient resistance. Among breast cancer detection methods, mammography is the most effective diagnostic method. For radiologists, diagnosing a cancerous mass on mammographic images is prone to error, which shows there is a need for a method to reduce the errors. In this study, a new adaptive thresholding method is proposed based on variable-sized windows. This method estimates the location of the mass and then determines the exact location of the cancerous tissue to reduce false positives. To detect the mass automatically, firstly, the histograms diagram and its relative maximums have been used to calculate the initial threshold for estimating the mass location. Two windows that contain information around each pixel and their size varies according to the mean value of each image due to the preservation of useful information. Secondly, two windows are used for the final threshold in order to discover the location of the mass and its exact shape. The proposed approach has been applied to 170 images of the Mammographic Image Analysis Society MiniMammographic database. Evaluations have shown 96.7% sensitivity and 0.79 false-positive rates, which prove an improvement in comparison with other state-of-the-art methods.

## 1 Introduction

Breast cancer is the most common form of cancer among women. Tumors of breast cancer are small at first and sometimes it takes several years to become a large gland, So early detection of tissue is ciritial. Breast cancer detection methods include: Breast self-examination, Breast examination by a doctor and Mammography, which the most effective way to detect breast cancer is Mammography. Incorrect detections in mammograpgy images by radiologists errors of analysing diagnosis due to physician fatigue or optical illusion, lead to the idea of finding a solution to do the process automatically. Radiologists fail to detect tissue in 10 to 30 percent of cases. In order to reduce these human mistakes, computer-aided detection/diagnosis (CAD) systems have been used [8]. Computer-aided detection systems are the most effective tools for early diagnosis of breast cancer. Segmentation is considered as one of the main steps in image processing. It divides digital images into multiple regions for better analysis. It can also specifies different objects in the image. Several segmentation techniques for image smoothing and easy evaluation have been developed by researchers. The most popular image segmentation techniques are: Edge Detection, Thresholding, Histogram, Region based methods and Watershed Transformation. Images are divided into two types based on their color (RGB images and Grayscale images). Therefore, the segmentation of RGB images and Grayscale images are completely different. Specifications of the pixels and information about their adjacent pixels are two fundamental parameters for any segmentation algorithm.

✉ Samaneh Mazaheri
Samaneh.Mazaheri@ontariotechu.ca

Bahram Sadeghi
b_sadeghi_b@iasbs.ac.ir

Meysam Karimi
meysam.k@iasbs.ac.air

[1] Department of Computer Science and Technology, Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, Iran

[2] Faculty of Business and Information Technology, Ontario Tech University, Oshawa, ON, Canada

One of the methods utilized to detect cancer tumor mass is morphological operators in the field of image processing and more specialized in the area of image segmentation. Morphological operators and Fuzzy C-means clustering used to isolate tissue or tumor mass in breast cancer images based on image processing method. Basha et al. [3], utilized Fuzzy C-means clustering, Morphological operator of erosion and dilation and found an appropriate answer to isolate the cancerous mass. In a research by Kekre et al., vector quantizatio method for image segmentation have been proposed that used a reasonable time but has high accuracy and low complexity [10]. Another method proposed by Mancatini et al. both the Microcalcification and cancer mass have been studied [15], also Moradkhani et al. proposed a new image mining approach for detecting micro-calcification in digital mammograms [16].

Thresholding is the simplest method of image segmentation. By using a threshold for a grayscale image, a binary or black-and-white image is obtained. In simplest method of thresholding each pixel is replaced by a pixel with a value of 0 or'black', if it is less than the fixed value T; otherwise if the value is greater than or equal to the value of the threshold or T, it is replaced by a value of 255 or white. In the method proposed by Zhang and Desai thresholds are chosen adaptively using wavelet and adaptive threshold. They are obtained by multi-objective analysis of Density Function Probability. This method does not apply to images that have a slight difference in the intensity of the cancer mass with the background [23]. Kom et al. presented an adaptive thresholding method to find a threshold for each pixel with respect to the adjacent pixels in a window. This method has been tested on 61 mammograms and has a sensitivity of 95.91%. Although this method has shown satisfactory result in exploring the mass, it fails to discover the center of the mass [12]. Following by that, a method was developed by Kai et al. to improve the defects of the Kom's method in which an adaptive threshold approach based on multiple analysis has been used. The results of the experiments show that the sensitivity of this method is 91.3% and the false positive rate is 0.97 on 170 MIAS database images [7]. In a method recently presented by Anita et al., mammograms are segmented in two stages; firstly a global threshold using the probability density function (PDF) and secondly, the final threshold using window-based thresholding for texture extraction. This method has a sensitivity of 93% and a false positive rate of 0.84 on two MIAS and DDSM databases [1]. Cancer tumors are usually classified in 6 categories: CIRC,[1] CALC,[2] SPIC,[3] MISC[4] ARCH,[5] and ASYM.[6] In this paper, besides CALC mass, others are discussed. The three main features

of the lesion include gray level, texture and shape in mammography. Two types of SPIC and ARCH mass are defined by the texture feature and Mass detection algorithms results in with the help of these features [2, 6, 9, 22]. CIRC masses are characterized in some algorithms by the shape attribute and have results in mass segmentation [5, 11, 14, 15]. Gray-level feature can well describe all types of masses, and detection algorithms with this feature report impressive performance. [1, 7, 12, 19, 23, 24].

Since early tissue detection is the best way to treat breast cancer, the main goal is to introduce a new method to detect a cancerous tumor by considering the following criteria:

- Automaticity of the proposed method
- Identify the exact location and shape of the mass
- High percentage of sensitivity in the diagnosis of cancerous mass
- Low false positive rate

This study has been utilized gray-level features, since they have a key role in mass detection algorithms, especially adaptive thresholding algorithms. This feature includes gray values that can be used to find the appropriate threshold for each of the two steps of the proposed method. In first step a threshold is obtained using the probability density function (PDF), then to reduce the false positive rate and detect the exact shape and location of the mass the final threshold based on variable-size windows is used. The results of this experiment show that the performance of the proposed method has significant improvements over other proposed methods. This paper is organized as follows: Sect. 2 covers the proposed method from the preprocessing stage to the final segmentation stage. In Sect. 3, the results of the implementation of the proposed algorithm will be described on mammography images along with several examples. Section 4 concludes the paper.

## 2 Methodology

Computer-Aided Detection Systems is a technology designed to reduce the amount of visual errors and false negative value derived from the physician's interpretation of medical images. In this paper a Computer-Aided Detection System is introduced which at first, it will remove images noise. Removing noise involves the removal of the pectoral muscle, removal horizontal and vertical lines. Removal of pectoral muscle is done by region growing algorithm with 2 or 3 seed points. The largest contour is known as the breast profile. Horizontal and vertical lines on

---

[1] Well-defined/circumscribed masses.

[2] Calcification.

[3] Spiculated masses.

[4] Other, ill-defined masses.

[5] Architectural distortion.

[6] Asymmetry.

the breast profile are removed by applying a median filter. To extract the breast profile, a threshold is defined for the image by applying the threshold and generating a binary image, the components of image will be extracted. After extraction of the breast profile, the global threshold is applied by the probability density function to estimate the position of the suspicious lesions. In the final stage, the cancer mass is extracted using a thresholding algorithm based on variable-size windows. Figure 1 shows steps of the proposed method.

## 2.1 Database

A public database is an essential step for efficient comparison and testing algorithm results. However, there are many issues with the establishment of a library of data which includes practical and theoretical problems. In the United Kingdom, the Society for the Analysis of Mammography Images decided to create a database of images with good and important criteria, but it has the following limitations:

1. The number of digitized images in the database was low due to time and financial constraints.
2. Mammographic images should only be collected from a location in the United Kingdom's national mammography screening program.
3. The images were taken from an angle of MLO only.
4. The time to collect clinical information was necessarily limited to a specific time period.
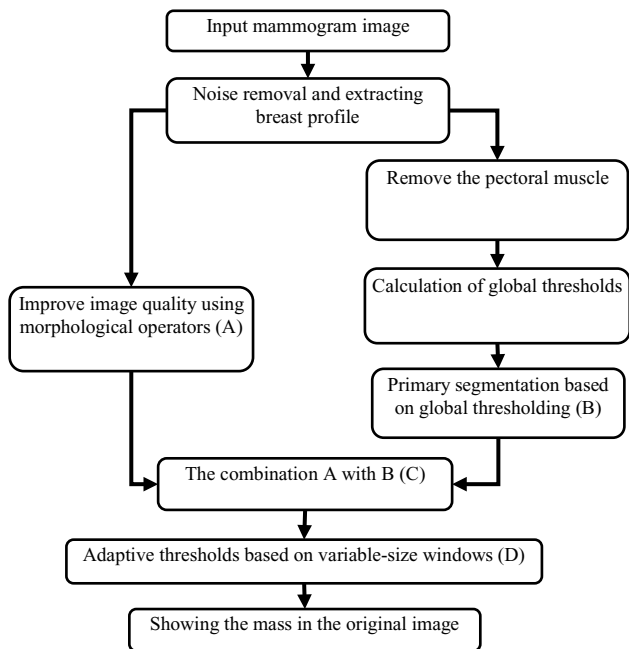


**Fig. 1** Flowchart; Steps of the proposed algorithm for the extraction of cancerous tissue

**Table 1** shows a summary of tissues types in the MIAS database.

| Film Category | Breast Type | | | TOTALS | |
|---|---|---|---|---|---|
| | Fatty | Fatty Glandular | Dense | | |
| Calcification | B: 2 | B: 5 | B: 5 | B: 12 | 25 |
| | M: 4 | M: 4 | M: 5 | M: 13 | |
| Circumscirbed Masses | B: 7 | B: 6 | B: 3 | B: 16 | 20 |
| | M: 2 | M: 2 | M: 0 | M: 4 | |
| Spiculated Masses | B: 2 | B: 4 | B: 6 | B: 12 | 21 |
| | M: 4 | M: 3 | M: 2 | M: 9 | |
| Architectural Distortions | B: 4 | B: 2 | B: 4 | B: 10 | 20 |
| | M: 2 | M: 4 | M: 4 | M: 10 | |
| Asymmetries | B: 2 | B: 3 | B: 3 | B: 8 | 17 |
| | M: 3 | M: 2 | M: 4 | M: 9 | |
| Miscellaneou | B: 4 | B: 3 | B: 1 | B: 8 | 15 |
| | M: 4 | M: 2 | M: 1 | M: 7 | |
| Normals | 66 | 64 | 74 | 204 | 204 |
| TOTALS | 106 | 104 | 112 | 322 | 322 |

Although the above criteria are not ideal, A database of images was created to collaborate on scientific research and CAD systems. The main mammography images from a Participant place in the National Mammography Imaging Program in England have been collected with high quality irradiation and suitable position of patient. (Table 1)

There is a list in this database's help file which includes characteristics of each image. Typically for an image with the following specifications:

*mdb*001   *G   CIRC   B*   535   425   197

will have
1. First column: MIAS database reference number
2. Second column: Character of background tissue

– F: Fatty
– G: Fatty-glandular
– D: Dense-glandular

3. Third column: Class of abnormality present:

– CALC: Calcification
– CIRC: Well-defined/circumscribed masses
– SPIC: Spiculated masses
– MISC: Other, ill-defined masses
– ARCH: Architectural distortions
– ASYM: Asymmerty
– NORM: Normal

4.Forth column: Severity of abnormality;

– B: Benign
– Malignant

**Table 2** Types of mass and number of images containing these masses in the MIAS database

| Type of mass | Number of image |
|---|---|
| CIRC | 22 |
| SPIC | 19 |
| ARCH | 19 |
| ASYM | 15 |
| MISC | 14 |
| TOTALS | 89 |

5. Fifth columns: x image-coordinate of centre of abnormality.

6 Sixth columns: y image-coordinate of centre of abnormality.

7. Seventh column: Approximate radius (in pixels) of a circle enclosing the abnormality [20].

Also in Table 2, the types of masses discussed in this paper, along with their number, are in the MIAS database.

## 2.2 Preprocessing

Preprocessing Mammography Images is the simplest diagnostic process for abnormalities without losing any useful data. Mammographic images may include noise during imaging. Also, parts of the breast area may be additional which is not necessary for the processing and analysis of these areas. Initial preprocessing is done on mammography images to eliminate noise and extract the breast profile. preprocessing steps used in this paper will be described in the following sections.

## 2.3 Removing horizontal and vertical lines

Mammographic images may include digitized noise such as straight lines which is visible in Fig. 2(b). These noises will be removed using the medianBlur() function (Median filter in OpenCV library) with the With a $3 \times 3$ core. The results of using this function in mammographic images are shown in Fig. 2.

## 2.4 Extracting the breast profile

The original mammography images in most databases contain physician's handwritten notes on their margins. In order to eliminate these handwritten as well as additional margins in mammographic images, first, a threshold of 20 is applied to convert the original image to a binary image. Figure 3 shows the disappearance of doctor's handwritten and the extraction of a breast profile.

## 2.5 Pectoral muscle removal

Pectoral muscle is a dense region close to the breast that is very effective in the accuracy of the texture detection algorithm. Given that intensity level of pectoral muscle in the image increase false positive rate in the results, so the removal of the muscle is very crucial. Due to using of the region growing algorithm to remove the pectoral muscle, it is necessary to determine the direction of the breast; to put the seed point in the proper position of the muscle. If the mammogram image is for the left breast the seed point is placed in the coordinates of the top left and if image is for the left breast the seed point is placed in the coordinates of the top right. In cases where the abnormality is joined to the chest muscle the first point is placed inside the chest muscle, the second point is placed inside the mass and the third point is inserted into the main part of chest. The image obtained from the removal of the pectoral muscle using three or two parent points and the region growing method is shown in Fig. 4. For images that have a tissue joined to the chest muscle, it is necessary to have three seed points, otherwise, two seed point is needed in order to remove the muscle.

**Fig. 2** Eliminate noise using the middle filter. a. Input mammogram image (mdb005). b. Part shown with digitized noise. c. Output image after clearing noise by applying median filter
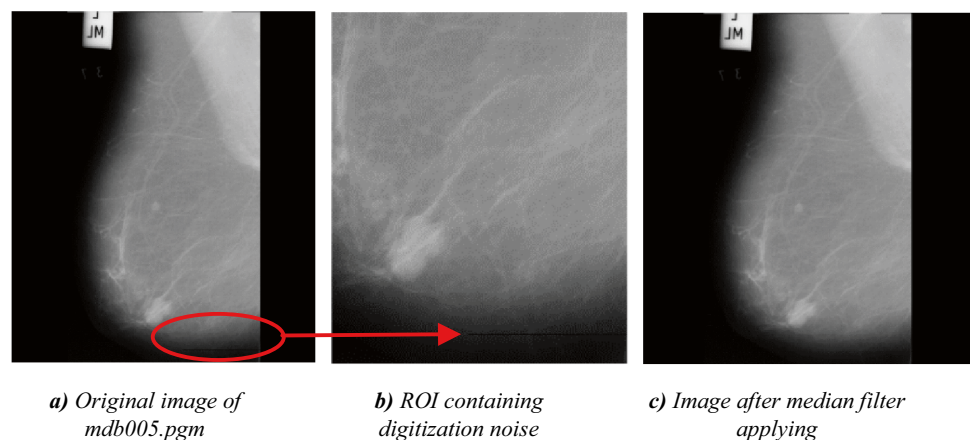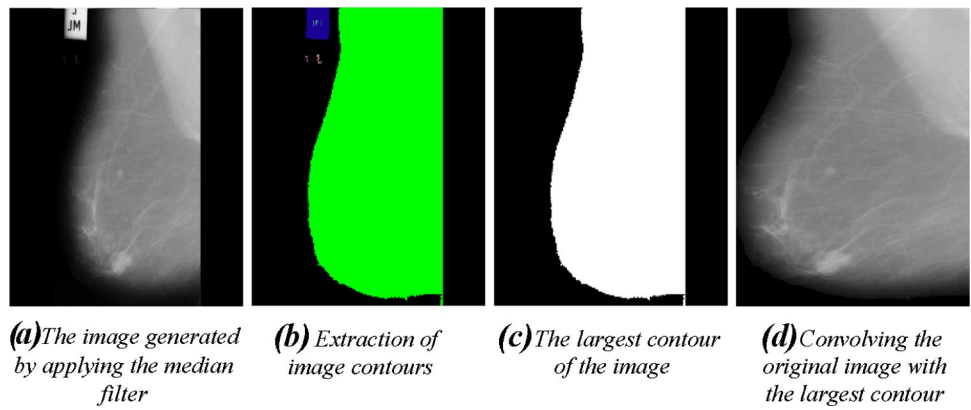


*a) Original image of mdb005.pgm*

*b) ROI containing digitization noise*

*c) Image after median filter applying*

**Fig. 3** Eliminating physician's handwritten and extracting breast profile

*(a)*The image generated by applying the median filter    *(b)* Extraction of image contours    *(c)*The largest contour of the image    *(d)*Convolving the original image with the largest contour

## 2.6 Aadaptive thresholding based on variable-size windows

In this paper, a method is presented to detect cancerous masses in two stages. The proposed method first finds the global threshold for the image using the probability density function (PDF), and after applying it, the image quality will be improved using morphological operators. The thresholding based on variable-size windows algorithm is used to determine the precise location and shape of the abnormal tissues. To improve the efficiency of the proposed method, it is necessary to combine both algorithms.

## 2.7 The first stage of the segmentation

Since the abnormalities in the mammography images have a higher gray level than the background image.

Differences in gray level distribution of tissue and background in histogram is high. The proposed method uses the histogram, such that a threshold for selection is more appropriate that the standard deviation between the background and the target is high on that threshold. First, the histogram for each mammography image is calculated, then the relative maximums are obtained from the histogram diagram. The relative maximums of histograms diagram are the candidates for choosing the global threshold for each image. Then the histogram will be normalized. This normalized histogram for the image with N pixels and L gray levels (L = 0, 1,…L-1) called PDF Which is obtained from Eq. (1).

$$PDF(i) = \frac{h(i)}{N} i = 0, 1, ...L - 1 \tag{1}$$

where $h(i)$ is the number of pixels with the gray level of $i$. Pixels with intensity less than 20 are ignored, since they
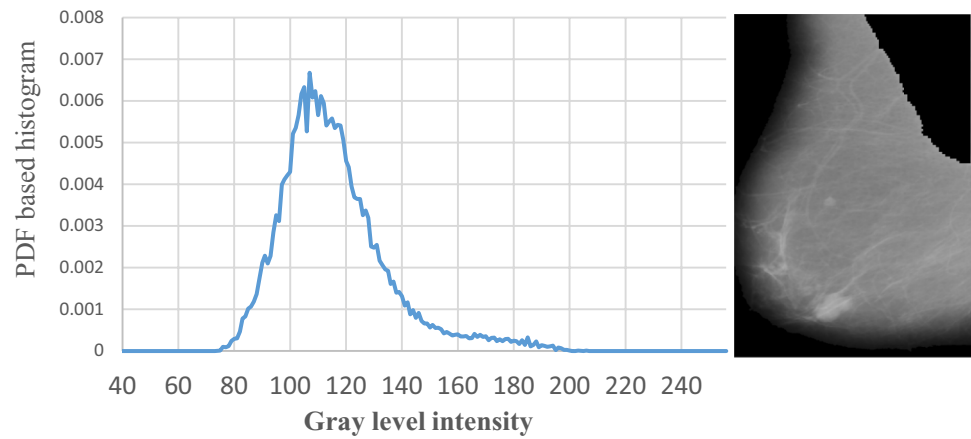


**Fig. 4** Removal of muscle using the region growing algorithm
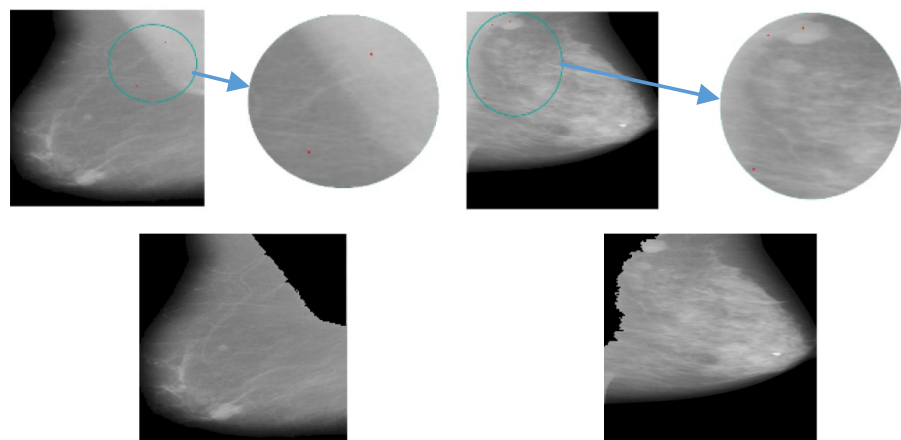
*(a) Remove pectoral muscle when muscle tissue is attached*

*(b) Removal of the pectoral muscle when the tissue is not attached to the muscle*

**Fig. 5** Histogram of preprocessed image



store background information. In the following, the global threshold calculate using the PDF obtained from Eq. (1) in Algorithm 1. In Fig. 5, the histogram of the preprocessed image is displayed.(Fig. 6)

In the second step of the Algorithm 1 $\tilde{x}$ is the mean $x_i$. In the third step of the Algorithm 1 $w_0(T)$ an $w_1(T)$ are calculated as the Eq. (2) and (3). Also, $S_0$ is the standard deviation of background pixels and $S_1$ is the standard deviation of the target pixels.

**Fig. 6** The combination of enhancement image and segmented image
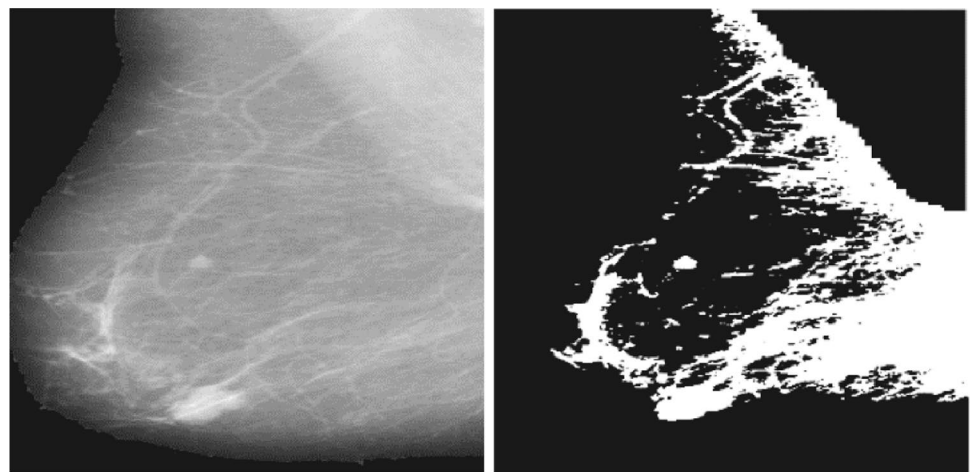


*(a)* Morphological enhanced image

*(b)* Segmented image with global threshold
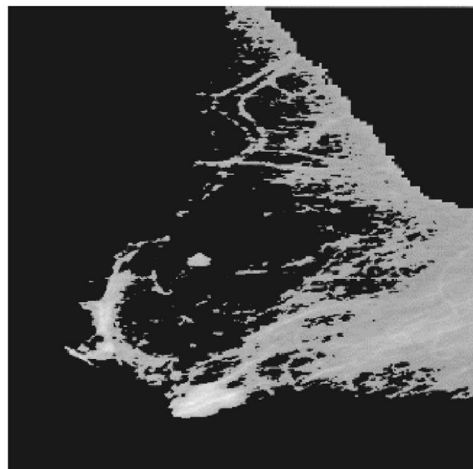
*(c)* Combined image

$$w_0(T) = \sum_{i=1}^{T-1} PDF(i) \tag{2}$$

$$w_1(T) = \sum_{i=T}^{L-1} PDF(i) \tag{3}$$

After obtaining the global threshold from relative maximums by using Eq. (4), the segmented image is resulted. Since the threshold ($T_{HPA}$) value above 200 causes the pixels to disappear and thus the cancer mass, so in this cases the threshold ($T_{HPA}$) is reduced to 180.

$$B(x,y) = \begin{cases} 1 & \text{if} A(x,y) > T_{HPA} \\ 0 & \text{if} A(x,y) \leq T_{HPA} \end{cases} \tag{4}$$

---

**Algorithm 1** Global Thresholding

---

1: Find the relative maximum of PDF

   Peak = (*i,PDF(i) | PDF(i)> PDF(i-1)& PDF(i)> PDF(i+1)*)

2: Calculate the standard deviation for each area

   with N pixels

   $S = \sqrt{\sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N}}$

3: Define the standard deviation between classes

   $\sigma_B^2(T) = w_0(T)w_1(T)[S_0(T) - S_1(T)]^2$

4: Calculate the global threshold

   $T_{HPA} = \max_{k \in Peak} \sigma_B^2(k)$

---

### 2.7.1 Utilizing morphological operators

Morphological operators can eliminate background and structural noise. The main task of these operators is to enhance the quality of the texture area of the image. At this stage, quality of preproposed images is significantly increased. This enhancement is performed with morphological operators. A top-hat operator has been implemented to eliminate the uneven background illumination.

$R(i,j) = A(i,j) - (A \circ s)(i,j)$

Such that $R(i,j)$ calculates the morphological opening of the image ($A \circ s$), with the structural element s and reducing it from the preprocessed image. *s* is a disc-shaped element with a radius of 60.

After enhancement pahse, the result is added to the preprocessed image.

$C(i,j) = (A(i,j) + R(i,j))$

In the final step, in order to improve the image quality, image *D* is obtained by combining image *B* with *C* as follows:

$$D(x,y) = \begin{cases} C(x,y) & \text{if} B(x,y) = 1 \\ 0 & \text{if} B(x,y) = 0 \end{cases} \tag{5}$$

### 2.7.2 Final segmentation

After combination of enhanced and segmented image, images enter the final stage of segmentation to extract the suspicious cancer tissue. At this stage, based on two windows with variable sizes, two threshold values are defined. Two windows with radius 10 and 30 are initially defined, the small window calculates the difference between the largest and smallest values around each pixel and the larger window, the average of the values around each pixel. Then the size of these two windows will be adjusted according to the mean value. The average of all mammographic images will be calculate and the difference between the mean value and the value of 200 is obtained. For every 30 differences, the window radius of one unit decreases. This decrementr in the radius of the smaller window is due to the fact that in low-average images, the possibility of mass loss increases. This may also occur in high-average images because of the less or equal intensity of cancer tumor in contrast to breast tissue.

$reduc = abs\ (200 - mean)/30$

The CLAHE[7] filters are used to enhance pixels in the images. Additionally, an opening morphological operator is used with images whose thresholds are greater than 160 at the thresholding of the initial segmentation step and their average values are high, also the window radius changes. For these images that have very high or very low mean value, the mean value is recalculated after applying the above mentioned operators and filters. After that two windows have been reached to the right

---

7 Contrast Limited Adaptive Histogram Equalization.

**Fig. 7** (a). Combination the image obtained from the primitive segmentation and the enhanced image (b). Result of the final segmentation (c). Extraction of contour that has a white value (d). Displaying the location of the mass with the red line
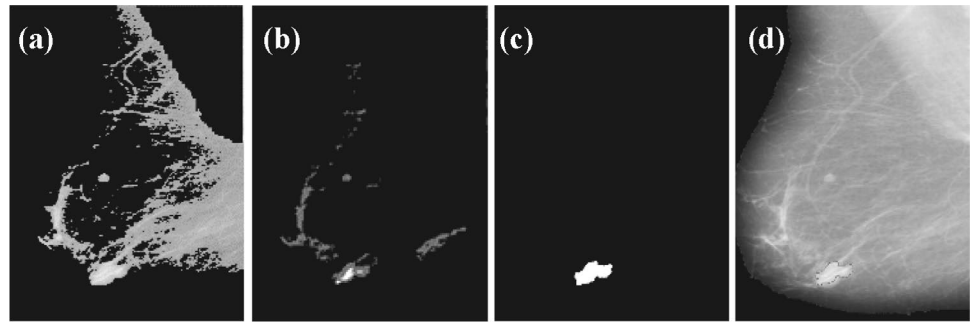


size. The quality of the images improved in accordance with the initial mean value. The Algorithm 2 is presented for calculating two thresholds. The first threshold is for appearing white color (value of 255 in third step) and second threshold is for appearing gray color (value of 100 in fifth) in images.

---
**Algorithm 2**  Variable-size Window-based Fine Thresholding

---
1: **if**  Wmean(x, y)) ≥  limite1 - 5 × $\alpha$  **then**

2:     If  D(x, y) ≥  (Wdiff(x, y) + limite2 - $\alpha$) and

       Count(x, y) > 200 **then**

3:     E(x,y) = 255

4:     else if  D(x, y) ≥  Wmean(x, y) **then**

5:     E(x,y) = 100

6:     **end if**

7: **end if**

---

The calculation of limite1 and limite2 is described in Eq. (6).

$$limite1 = \frac{(THC1+mean1)}{2}$$
$$limite2 = \frac{(THC1+THC1)}{2} \tag{6}$$

In algorithms 3 and 4, the calculation of THC1 and THC2 is expressed. *Wdiff*, a small window corresponding to each pixel, is obtained using the Eq. 7.

$$Wdiff(x, y) = Wmin(x, y) - Wmax(x, y) \tag{7}$$

In Eq. (7) *Wmin*(x, y) and *Wmax*(x, y) are the minimum and maximum values in the small window for each pixel, respectively. Also, the *Wmean*(x, y) mentioned in the Algorithm 1 is the mean of the large window associated with each pixel.

---
**Algorithm 3** THC1

---
1: **if**   Initial_T > 190  **then**

2:     THC1 = 190

3: **else if**  mean1 > 195  **then**

4:     THC1 = mean1

5: **else**

6:     THC1= T2 > mean2 ? mean2 : T2

7: **end if**

---

So that Initial_T is related to the initial threshold in the initial segmentation step. Also, the value of T2 is obtained using the initial threshold when the background value of the image is increased to 190. *mean*1 is the average of the images entered to the final stage. *mean*2 is the average of all pixels whose values are greater than *mean*1.

---
**Algorithm 4** THC2

---
1: **if**   mean1 < 150 **then**

2:     THC2 =  mean1

3: **else if**  mean1 > 195  **then**

4:     THC2 = $\frac{(mean1+THC1)}{2}$

5: **else**

6:     THC2= mean1

7: **end if**

---

Using algorithm 2, suspicious cancer cells are extracted. If no mass is found, the algorithm repeats until the mass is observed. This is done by resizing two windows.

**Table 3** Comparison of sensitivity for the proposed method in the segmentation of all types of masses with other works

| Type of mass | Number (masses / images) | Kai's method [7] | | Anita's method [1] | | Proposed method | |
|---|---|---|---|---|---|---|---|
| | | Number of masses | sensitivity (%) | Number of masses | sensitivity (%) | Number of masses | sensitivity (%) |
| CIRC | 24/24 | 23 | 95.8 | 24 | 100 | 24 | 100 |
| SPIC | 19/19 | 15 | 78.9 | 16 | 84.9 | 18 | 94.7 |
| ARCH | 19/19 | 18 | 94.7 | 17 | 89.5 | 18 | 94.7 |
| ASYM | 15/15 | 14 | 93.3 | 14 | 93.3 | 15 | 100 |
| MISC | 14/15 | 14 | 93.3 | 15 | 100 | 14 | 93.3 |
| TOTAL | 89/92 | 84 | 91.3 | 86 | 93.5 | 89 | 96.7 |

After final segmentation step, according to the algorithm 2, two groups of contours are shown with white and gray color in the image. The gray contours contained in the white contours represent the shape and location of the cancerous tissue. The steps of mass discovery in the final segmentation are shown in the Fig. 7.

## 3 Results of image segmentation

In this implementation, an automatic segmentation algorithm is performed on mammographic images. The efficient criteria of the algorithm to evaluate the segmentation results, include sensitivity and number of false positives which are calculated using the Eq. 8 and Eq. 9.

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$FP/image = \frac{Number\ of\ FP\ identified}{Total\ number\ of\ images} \tag{9}$$

These criteria are calculated based on the region-based evaluation. This means that an area is known as a true positive (TP), if the percentage of overlap for the mass of the specified area by the radiologist with the computer-segmented area is greater than 50%. The false negative rate is the amount of mass obtained by the radiologist, which is not recognized by the proposed method. False Positive Error is the number of masses that mistakenly detected and not detected by the radiologist. The results of the detection of suspicious masses for various abnormalities are shown in the Table 3. This table compares the sensitivity of different types of cancerous masses. The comparison was done between the proposed method and the method described by Hu et al. [7] as well as Anita et al. [1]. The proposed algorithm has a sensitivity of 100% in the type of ASYM and CIRC anomalies.

Improvement in image quality by morphological operators is effective to detect masses with a CIRC type. The size of the two windows are also changed in order to locate the masses of this type. For masses that are less intense, the morphological operations of opening with an oval structure element makes the mass more clear. An ellipsoidal structural element in morphological operators is suitable for circular masses. In the case of mass MISC type because of the higher brightness of the mass in compare with background, this algorithm performs better. Also, the mass of the ASYM type

**Table 4** Comparison of the effectiveness of the proposed method with other methods for the segmentation of MIAS images

| Method | Sensitivity (%) | False positive rate |
|---|---|---|
| Spatial clustering [4] | 88.7 | 3.46 |
| Adaptive thresholding [7] | 91.3 | 0.97 |
| Multiple thresholding (wavelet transform and genetic algorithm) [17] | 90.0 | 2.63 |
| Havrda and Charvat entropy and Otsu thresholding [13] | 90.2 | 8.21 |
| Cellular neural network [18] | 90.9 | 5.30 |
| Wavelet processing and adaptive thresholding [21] | 91.0 | 2.23 |
| Two-step adaptive thresholding [1] | 93.5 | 0.84 |
| Adaptive thresholding based on variable-size windows (proposed method) | 96.7 | 0.79 |
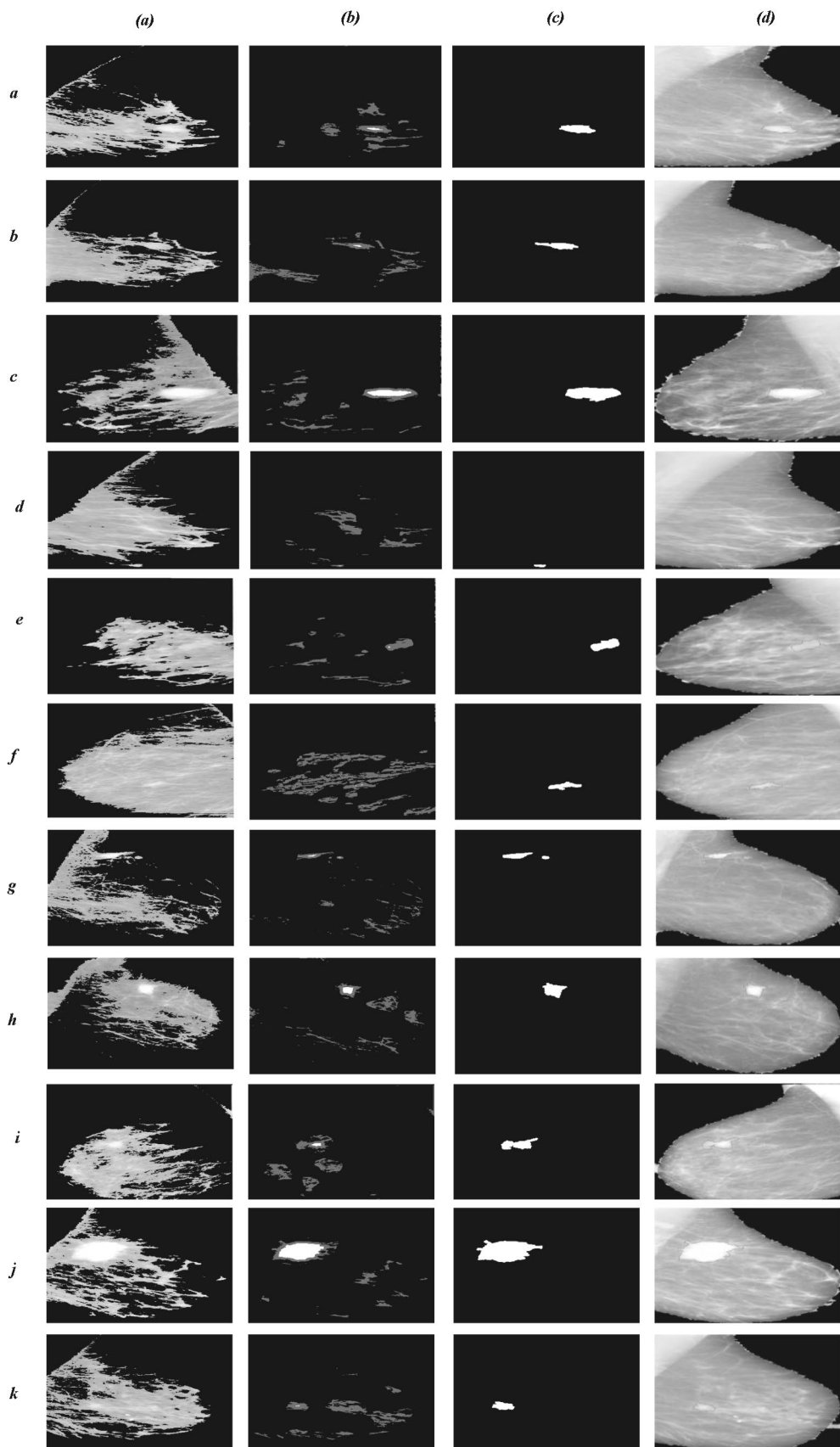
**◄Fig. 8** Column (a). Combination the image obtained from primitive segmentation and enhanced image. Column (b). The first part of the adaptive segmentation based on variable-size windows. Column (c). Mass extraction using proposed algorithm. Column (d). Displays the mass with the red line in enhanced image. Row (A): The *mdb*010 image is from the *MIAS* database. Row (B): The *mdb*012 image is from the *MIAS* database. Row (C): The *mdb*025 image is from the *MIAS* database. Row (D): The *mdb*080 image is from the *MIAS* database. Row (E): The *mdb*091 image is from the *MIAS* database. Row (F): The *mdb*097 image is from the *MIAS* database. Row (G): The *mdb*132 image is from the *MIAS* database. Row (H): The *mdb*134 image is from the *MIAS* database. Row (I): The *mdb*155 image is from the *MIAS* database. Row (J): The *mdb*184 image is from the *MIAS* database. Row (K): The *mdb*204 image is from the *MIAS* database.

is easily recognizable due to its large size and high mass density. In the case of the two SPIC and ARCH categories, due to the needle-shaped structure, it is a bit difficult to discover the location of the mass because the proposed algorithm does not detect the mass by texture features, but it extracts the cancerous mass using properties related to light intensity and shape of texture. In the Fig. 8, the results of the proposed algorithm implementation in the *MicrosoftVisualStudio* environment are shown on several images of the MIAS database.

The results in Table 4 indicate that the proposed method has better performance in two criteria of sensitivity and false positive rate than other methods.

Table 5 compares the result of proposed method with two previous works in this field. This comparison is preformed on 54 images. The number of masses discovered for each image is shown by three methods. All three algorithms are implemented in images from the same database called MIAS.

The precision of the proposed method in the segmentation of mammography images can help researchers in the problem of features extraction and classifications.

**Table 5** Comparison of the proposed method and two other methods for detecting the number of masses in 54 images

| Image name | Type of mass | Proposed method | Kai method [7] | Anita method [1] | Number of masses | Image name | Type of mass | Proposed method | Kai method [7] | Anita method [1] | Number of masses |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Mdb001 | CIRC | 1 | 1 | 1 | 1 | Mdb110 | ASYM | 3 | 1 | 2 | 1 |
| Mdb002 | CIRC | 2 | 1 | 1 | 1 | Mdb111 | ASYM | 3 | 1 | 1 | 1 |
| Mdb015 | CIRC | 2 | 4 | 1 | 1 | Mdb115 | ARCH | 1 | 1 | 1 | 1 |
| Mdb017 | CIRC | 3 | 5 | 3 | 1 | Mdb117 | ARCH | 2 | 1 | 1 | 1 |
| Mdb019 | CIRC | 2 | 2 | 2 | 1 | Mdb120 | ARCH | 1 | 2 | 2 | 1 |
| Mdb021 | CIRC | 2 | 1 | 1 | 1 | Mdb121 | ARCH | 1 | 1 | 1 | 1 |
| Mdb023 | CIRC | 2 | 4 | 1 | 1 | Mdb124 | ARCH | 3 | 1 | 1 | 1 |
| Mdb244 | CIRC | 2 | 1 | 1 | 1 | Mdb125 | ARCH | 3 | 2 | 3 | 1 |
| Mdb270 | CIRC | 2 | 2 | 1 | 1 | Mdb127 | ARCH | 2 | 3 | 2 | 1 |
| Mdb290 | CIRC | 2 | 3 | 2 | 1 | Mdb163 | ARCH | 3 | 1 | 1 | 1 |
| Mdb315 | MISC | 2 | 1 | 1 | 1 | Mdb165 | ARCH | 1 | 1 | 1 | 1 |
| Mdb013 | MISC | 1 | 2 | 1 | 1 | Mdb170 | ARCH | 2 | 2 | 3 | 1 |
| Mdb030 | MISC | 3 | 6 | 3 | 1 | Mdb171 | ARCH | 1 | 3 | 3 | 1 |
| Mdb032 | MISC | 3 | 2 | 3 | 1 | Mdb145 | SPIC | 1 | 2 | 3 | 1 |
| Mdb058 | MISC | 1 | 2 | 1 | 1 | Mdb175 | SPIC | 2 | 3 | 1 | 1 |
| Mdb063 | MISC | 3 | 2 | 1 | 1 | Mdb178 | SPIC | 1 | 1 | 1 | 1 |
| Mdb264 | MISC | 1 | 1 | 1 | 1 | Mdb179 | SPIC | 1 | 1 | 1 | 1 |
| Mdb265 | MISC | 2 | 1 | 1 | 1 | Mdb181 | SPIC | 2 | 1 | 1 | 1 |
| Mdb072 | ASYM | 3 | 1 | 1 | 1 | Mdb186 | SPIC | 2 | 1 | 1 | 1 |
| Mdb081 | ASYM | 1 | 1 | 1 | 1 | Mdb188 | SPIC | 1 | 1 | 2 | 1 |
| Mdb083 | ASYM | 2 | 1 | 1 | 1 | Mdb190 | SPIC | 4 | 2 | 3 | 1 |
| Mdb090 | ASYM | 3 | 2 | 2 | 1 | Mdb191 | SPIC | 1 | 5 | 4 | 1 |
| Mdb099 | ASYM | 2 | 5 | 3 | 1 | Mdb193 | SPIC | 2 | 1 | 1 | 1 |
| Mdb102 | ASYM | 2 | 3 | 1 | 1 | Mdb198 | SPIC | 3 | 1 | 1 | 1 |
| Mdb104 | ASYM | 1 | 2 | 1 | 1 | Mdb199 | SPIC | 2 | 3 | 1 | 1 |
| Mdb105 | ASYM | 1 | 1 | 1 | 1 | Mdb202 | SPIC | 3 | 1 | 1 | 1 |
| Mdb107 | ASYM | 3 | 2 | 4 | 1 | Mdb207 | SPIC | 1 | 1 | 2 | 1 |

# 4 Conclusion

In this paper, a method is presented in which the images are first preprocessed to produce a satisfactory result for the segmentation phase. At this stage, the median filter for the removal of Linear shaped noise, also used the Region growing method using three seed points to remove the pectoral muscle. Then, the largest contour was extracted as a breast profile. After the preprocessing stage, the initial segmentation is performed to estimate the location of the mass. Next the image quality enhanced. In the second stage of segmentation, an adaptive thresholding algorithm is used. This algorithm is based on variable-size windows that can extract cancer mass. The proposed method is implemented on the MIAS database images and its accuracy in mass extraction is 96.7% (89 images out of 92 images). This method has increased sensitivity criteria compared to the method presented by Anita [1] 3.7%. Also, the false positive rate has decreased by 0.05. The proposed method improved by 4.9% compared to the Kai method [7] in the sensitivity.

and False-positive rate decreased amount of 0.18. The precision of the proposed method in the segmentation of mammography images can help researchers in the problem of features extraction and classifications. In the future, to test the proposed algorithm, it will be implemented on other databases. A classification method will also be provided to determine which category of extracted mass (malignant, benign and normal).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no known competing financialinterests or personal relationshipsthat could have appeared to influence the work reported in this paper.

# References

1. Anitha J, Peter JD, Pandian SI. A dual stage adaptive thresholding dusat for automatic mass detection in mammograms. Comput Methods Programs Biomed. 2017;138:93–104.
2. Ayres FJ, Rangayyan RM. Characterization of architectural distortion in mammograms. IEEE Eng Med Biol Mag. 2005;24(1):59–67.
3. Basha SS, Prasad KS. Automatic detection of breast cancer mass in mammograms using morphological operators and fuzzy c–means clustering. Journal of Theoretical & Applied Information Technology. 2009;5(6)
4. Cao A, Song Q, Yang X. Robust information clustering incorporating spatial information for breast mass detection in digitized mammograms. Comput Vis Image Underst. 2008;109(1):86–96.
5. Groshong BR, Kegelmeyer WP. Evaluation of a hough transform method for circumscribed lesion detection. Digital mammography. 1996;96:361–6.
6. Guliato D, Rangayyan RM, Carvalho JD, Santiago SA. Polygonal modeling of contours of breast tumors with the preservation of spicules. IEEE Trans Biomed Eng. 2008;55(1):14–20.
7. Kai Hu, Gao X, Li F. Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms. IEEE Trans Instrum Meas. 2011;60(2):462–72.
8. Jalalian A, Mashohor SB, Mahmud HR, Saripan MI, Ramli AR, Karasfi B. Computer aided detection diagnosis of breast cancer in mammography and ultrasound a review. Clin Imaging. 2013;37(3):420–6.
9. Karssemeijer N, te Brake GM. Detection of stellate distortions in mammograms. IEEE Transactions on Medical Imaging. 1996;15(5):611–9
10. Kekre HB, Sarode TK, Gharge SM. Tumor detection in mammography images using vector quantization technique. International Journal of Intelligent Information Technology Application. 2009;2(5):237–42.
11. Kobatake H, Murakami M, Takeo H, Nawano S. Computerized detection of malignant tumors on digital mammograms. IEEE Trans Med Imaging. 1999;18(5):369–78.
12. Kom G, Tiedeu A, Kom M. Automated detection of masses in mammograms by local adaptive thresholding. Comput Biol Med. 2007;37(1):37–48.
13. Kurt B, Nabiyev VV, Turhan K. A novel automatic suspicious mass regions identification using havrda & charvat entropy and otsu's n thresholding. Comput Methods Programs Biomed. 2014;114(3):349–60.
14. Li H, Wang Y, Liu KR, Lo SC, Freedman MT. Computerized radiographic mass detection I lesion site selection by morphological enhancement and contextual segmentation. IEEE Trans Med Imaging. 2001;20(4):289–301.
15. Mencattini A, Salmeri M, Lojacono R, Frigerio M, Caselli F. Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing. IEEE Trans Instrum Meas. 2008;57(7):1422–30.
16. Farzaneh Moradkhani and Bahram Sadeghi Bigham. A new image mining approach for detecting micro-calcification in digital mammograms. Applied Artificial Intelligence. 2017;31(5–6):411–24.
17. Pereira DC, Ramos RP, Nascimento MZD. Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm. Comput Methods Programs Biomed. 2014;114(1):88–101.
18. Wener Borges Sampaio, Edgar Moraes Diniz, Aristófanes Corrêa Silva, Anselmo Cardoso De Paiva, and Marcelo Gattass. Detection of masses in mammogram images using cnn, geostatistic functions and svm. Computers in Biology and Medicine. 2011;41(8):653–64
19. Singh S, Bovis K. An evaluation of contrast enhancement techniques for mammographic breast masses. IEEE Trans Inf Technol Biomed. 2005;9(1):109–19.
20. John Suckling, J Parker, D Dance, S Astley, I Hutt, C Boggis, I Ricketts, E Stamatakis, N Cerneaz, S Kok, et al. The mammographic image analysis society digital mammogram database. In Exerpta Medica. International Congress Series. 1994;375–8
21. Vikhe PS, Thool VR. Mass detection in mammographic images using wavelet processing and adaptive threshold technique. J Med Syst. 2016;40(4):82.
22. Zhang M, Giger ML, Vyborny CJ, Doi K. Mammographic texture analysis for the detection of speculated lesions. Excerpta Medica. 1996;1119:347–51.
23. Zhang XP, Desai MD. Segmentation of bright targets using wavelets and adaptive thresholding. IEEE Trans Image Process. 2001;10(7):1020–30.
24. Xiao-ping Zhang. Multiscale tumor detection and segmentation in mammograms. In Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium. 2002;213–6