



AUTOMATIC IMAGE SEGMENTATION METHOD FOR BREAST CANCER ANALYSIS USING THERMOGRAPHY

Aya Hossam¹, Hany M. Harb², and Hala M. Abd El Kader³

^{1,3} *Electrical Eng. Depart, Faculty of Engineering (Shoubra), Benha University, Cairo, Egypt*

² *Computers and Systems Eng. Depart, Faculty of Engineering, Azhar university, Cairo, Egypt*

Received 16 October 2017; Accepted 4 November 2017

ABSTRACT

Breast Cancer (BC) is considered one from various diseases that has got great attention in the last decades. This is because the high mortality rates among young women in the whole world according to this disease. The early detection of this disease is necessary to enhance the opportunity of survival. Thermography is an efficient screening tool that can help in detecting BC by indicating parts of the body where an abnormal temperature variation is found. To realize an effective BC detection system using thermography, it is necessary to locate the region of interest (ROI) in the thermograms prior to analysis. This paper introduces a new automatic segmentation method (SM) for identifying the ROI image from breast thermograms. It depends on the statistics of image in DMR-IR database. After detecting the boundaries of ROI using this new SM, two different approaches for accurate identifying and enhancing the breast boundaries are investigated: the Hough transform (HT) algorithm to locate and define the parabola curves in the breast image, and a second approach introducing the enhancement of the contrast of detected boundaries in the breast ROI image. To evaluate the results of this method, statistical features have been extracted from the segmented ROI image. Then, we used the Support Vector Machine (SVM) and the Artificial Neural Network (ANN) classifiers to detect the normal and abnormal breasts based on these features. The experimental results prove that the accuracy of SVM, and ANN classifiers reached to 96.67% and 96.07%, respectively. This denotes that the proposed automatic SM is a favorable technique for extracting the breast ROI image from breast thermograms.

Keywords: Breast Cancer, Thermography, Region of Interest, Segmentation Method, Classifiers.

1. Introduction

BC became the most generic form of cancer between women in worldwide which can cause death [1],[2]. Every year, more than 20% of newly cancer cases are diagnosed by BC. The danger of being diagnosed with BC increases with age. In Canada, it is evaluated that 83% of the new BC cases occurred among women over the 50 years old in 2016 [3]. In USA, breast cancer is the most commonly diagnosed cancer among American women. In 2017, it's estimated that about 30% of newly diagnosed cancers in women will be breast cancers according to American Cancer Society [4]. In Egypt, BC is the most common cancer among women and is evaluated to be the cause of 22% all cancer-related women deaths [5]. Early detection and fast diagnosis of BC is the key to save the patient's life and decrease mortality rate. The opportunities of recovery would be increased by 85%, if the

BC is detected at early stages. However, this percentage is reduced to 10%, when the disease is found in advanced stages [6].

A recent study estimates the performance of various breast imaging techniques for early detection of BC [7]. Mammography is known as the gold standard imaging technique for detecting BC. It uses low energy X-ray radiation that produces image of breast area on a high contrast and high-resolution film [8],[9]. However, this procedure has some limitation [10], [11]. Firstly, it is painful for the patient according to the compression of the breast. Besides that, the X-ray radiation exposure may damage the cancerous tissue and can cause mutation. The breast density also affects the sensitivity of mammogram images. Finally, a lot of searches detect that mammogram is less effective for women less than 50 years old [11], [12].

According to these limitations of mammography procedure, a high concern has been given to thermography as an effective BC screening tool. As known for a temperature above absolute zero, all bodies emit infrared (IR) radiation. Breast thermography procedure uses a thermal infrared camera which converts this IR radiation into electrical signals, and presents them as a thermogram. So, the dangerous tissue with cancer is highlighted and easily distinguished from a normal tissue in a thermogram [13], [14]. Breast thermography has some important advantages over mammography such as its ability to deal with dense breast tissues, effective in all age groups and it also can easily be used for men [15]. Thermography is also harmless, fast method, and helps in early detection of BC [14]. Finally, the ionization, high pressure, and compression of breast are not required in thermography so no rupture risk here [16], [17].

The extraction of the Region of Interest (i.e., ROI segmentation) from the thermogram is a very important step in any BC detection system. ROI segmentation is used to extract and separate the breast regions from the other parts of the body in the thermogram. It can range from a completely manual to a fully automatic process. The features extraction process can be done, once the breast region is separated from the rest of breast thermogram.

There are several papers in breast thermogram segmentation search area. In [18], the separation process of breast parts from the overall thermal images was manually prior to analysis. It presents a segmentation protocol at which each breast is split into four quadrants. The breasts are divided by four reference points: the chin, the left side, the right side and the bottom edge of the breast. The image is separated into four different quadrants by these four points which connected to the nipple. For the various automatic SM, as in [19],[20], the level set technique [21] was presented to extract the blood vessels in a thermogram. In [19], an automatic SM is introduced using active contour and level set method without re-initialization. It is used to separate the breast regions from breast thermograms. This paper used a statistical based noise removal and contrast-limited-adaptive histogram equalization to enhance signal to noise ratio and the contrast of thermograms before applying the level set. In [22], the segmentation of breast ROI is done by extracting the body boundary using Canny edge detector and removal of inner edges. Then, it uses Fuzzy c-means clustering approach for segmentation of the breast part. After that, Radon projection is applied to these segmented images for the bi-spectral invariant features extraction. In [23], a SM is proposed at which the background region is extracted by using the Otsu's thresholding method and after that a reconstruction technique is used. Then the inframammary tuck is detected to locate the breast lower limit. Then, the upper limit of the breast is located by discerning the axilla. Based on these two limit lines, the breast part is separated finally.

The image preprocessing and segmentation steps are the essential requirements for the relevant features extraction process from the breast thermograms. Development of a fully automatic SM is a very difficult task because of their amorphous nature and unobvious limits. It is also based on the distance from which the breast thermogram is captured, the image height, the breast size, image background, presence of noise etc.. There are several attempts [24], [25],[26] which were made to demonstrate the ability of distinctive features in detecting BC abnormalities. Texture features are used to detect abnormal thermograms through using classification algorithms like SVM and ANN. Statistical features range from first order statistical to higher order statistical features and they can be extracted from the separated ROI images to analyze the abnormality of these images.

In this paper, an automatic breast SM is introduced based on the breast thermogram statistics of DMR-IR database. The thermal images at first must pass through some steps of pre-processing such as filtering images from noise, converting RGB images to grayscale images etc. Then, the proposed SM is used which provides an appropriate technique to separate the left and right breast parts for analysis. After separating the ROI image from the thermal image, different selective texture features are extracted from these segmented breast images. To evaluate this new SM, the extracted features are given as input to a two classifiers which are SVM, and ANN. The experimental results prove that the proposed SM improves the performance of classifiers with high accuracy values.

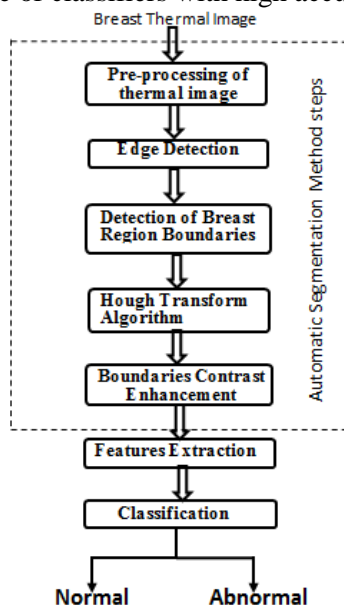


Fig. 1. Flow chart of BC Detection Model Steps.

2. The proposed SM

In this section, a detailed method for automatic segmentation of BC thermograms is presented. In this paper, breast thermograms are obtained from an open online data base PROENG (<http://visual.ic.uff.br/>) which called DMR-IR database [27]. Total of 200 (90 normal and 110 abnormal) thermal images were considered for this work. The thermograms pass through many steps due to the proposed automatic SM. It consists of mainly two steps: pre-processing, and segmentation. The segmentation step is divided into subparts which are ROI detection and separation, segmentation of ROI image into left and

right parts of breast, applying HT algorithm for accurate identifying the bottom parabola breast boundaries, and finally improve the contrast of detected boundaries in the selected breast region. After finishing segmentation step, features can be extracted from the segmented breast ROI images. For evaluation of the proposed SM, the significant extracted features are given as inputs to a SVM and ANN classifiers. Figure-1 shows a flow chart of a BC detection model involving the steps of our new SM.

2.1. Pre-processing

The captured breast thermograms are obtained using the FLIR SC-620 thermal IR camera. This camera converts the IR radiation that emitted from the skin surface into electrical signals and the output body image is visualized in color on a monitor. The captured thermogram is 640x480 pixels, that has a resolution of 24-bit per pixel. The original breast thermograms include some labels and signs according to the vendor software, as shown in [Figure-2a]. Thus, removing those labels and signs is the first thing must be done in the preprocessing step of the proposed SM as shown in [Figure-2b]. Next, it is converted into 8-bit grayscale image. All those labels and signs are removed using dedicated vendor software. Finally, using National Television System Committee (NTSC) conversion formula (equation 1) to convert the 24-bit color images into a corresponding 8-bit grayscale image. The result grayscale image is shown in [Figure-2c].

$$G_{gray} = (0.299 * red_{component}) + (0.587 * green_{component}) + (0.114 * blue_{component}) \quad (1)$$

where G_{gray} is the 8-bit gray scale image.

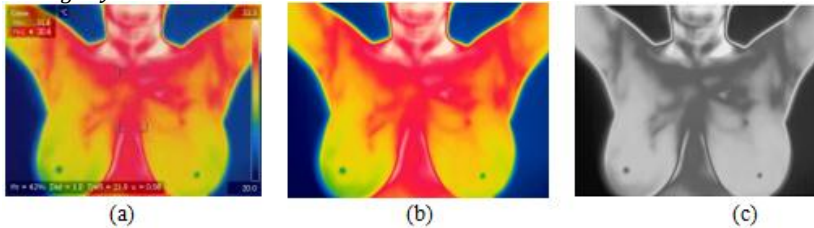


Fig. 2. (a) The original breast thermogram, (b) after processing using FUR software, (c) grayscale image.

2.2. Edge detection

Edge detection helps in capturing the objects due to their closed outline in the image [28]. In this step, both outlines and boundaries between objects and the background in the thermogram are detected. It is implemented here using the Canny detector due to its robustness to noise and its equal treatment of false positives and false negatives [29]. The Canny edge detector is a multi-step method that used to detect edges in an image. This is done by looking for local maximum values of the gradient of $f(x, y)$. The gradient is evaluated by doing the derivative of a Gaussian filter to obtain the magnitude and orientation of the gradient for each pixel [30]. This method uses two thresholds to identify strong and weak edges. The weak edges are involved in the output only if they are connected to strong edges. To get much smooth image and delete the unwanted edges, we must change the values of these thresholds. This paper uses the built-in Canny edge detector in the MATLAB Image Processing Toolbox. In [Figure-3], the results of this step are presented.

2.3. Detection of breast region boundaries

In this section, there are four important limit lines must be detected to isolate the breast part only from the thermal image. These lines are the upper, lower, left, and right limit

lines. In the used BC thermograms, the distance between body and camera is around 1 meter (dynamic protocol) and also the breast's region occupies about half of the thermogram height while the other areas include women shoulders and stomach as in [31].

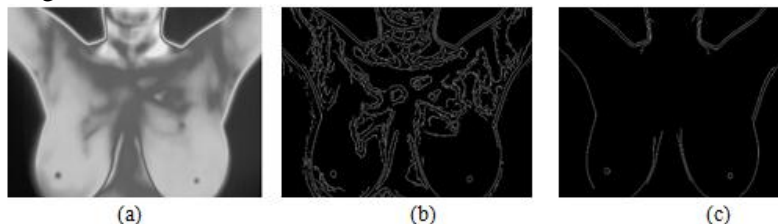


Fig. 3. (a) Original grayscale image, (b) Edge detection using Canny method without thresholding, (c) Edge detection using Canny method identifying threshold values.

The algorithm that has been implemented to detect the upper, lower, left, and right limit lines is summarized as in Algorithm (1).

Algorithm (1) Breast Region Boundaries Detection.

1. Load the original grayscale image, G .
 2. Apply Canny Edge Detector.
 3. Read $H = G$'s height.
 4. Read $W = G$'s width.
 5. Read the coordinates Y_1 and Y_2 where $Y_1 = \frac{3}{5} * H$ and $Y_2 =$ the minimum value of Y coordinate on breast edge.
 6. Read the coordinates X_1 and X_2 where $X_1 =$ the min. value of X coordinate on the breast edge and $X_2 =$ the max. value of X coordinate on the breast edge.
 7. Separate the breast ROI where $ROI = imcrop(G, [X_1, X_2, Y_1, Y_2])$.
 8. Determine the central axis value to segment the breast ROI image as left and right breast images.
-

2.3.1. Upper and lower limit lines

After reading the breast thermograms height in [27], we found that the upper limit (UL) line can be detected at about 0.6 of the total height of the image. So, if the image height is H , we can find the UL line(Y_1) by the following equation:

$$Y_1 = \frac{3}{5} * H \quad (2)$$

The next step is to identify the lower limit (LL) line(Y_2). In this step, the proposed SM depends on tracking the body outline edge after applying the Canny edge detector on the grayscale image. Firstly, SM searches on the body outline edge for the first minimum pixel location in the thermograms, and to find the least minimum value, more than 10 pixels are taken and compared together. Then, continue with the body outline edge until reaching the edge end of the breast in thermal image. At this point the minimum value of Y coordinate can be found and the LL line will cut this point (Y_2). Figure-4a shows the position of UL and LL lines. The output segmented image is shown in [Figure-4b].

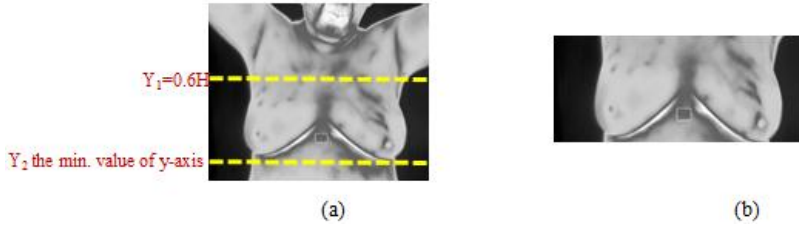


Fig. 4. (a) The detected position of upper and lower limit lines, (b) The output segmented image.

2.3.2. Left and right limit lines

Now, breast thermogram is separated from the unwanted top and bottom parts, by using the UL and LL lines. Left limit (FL) and right limit (RL) lines are detected from the segmented breast image by also tracking the body outline edge to find the minimum (X_1) and maximum (X_2) points. The FL and RL lines must cut these minimum and maximum points, respectively. The steps followed for the localization of FL and RL limit lines are given below:

1. After extracting the UL and LL lines, the proposed SM will track the body outline edge to evaluate the minimum point value of X coordinate on the breast edge, which symbolised by X_1 , using the coordinate matrix of this image.
2. The FL line is that cut this point (X_1) as shown in [Figure-5a].
3. By tracking the boundary of segmented image, we can also find the maximum point (X_2) of the outer breast edge on the X coordinate. The RL line will cut this point as shown in [Figure-5a]. In [Figure-5b], the output ROI after defining the upper, lower, left, and right limit lines are shown.



Fig. 5. (a) The detected position of FL and RL lines, (b) The output of breast region only.

The detected FL and RL lines are important to form a desired thermogram of the breast by removing the undesired left and right parts of the image. Now, the breast should be split into left and right parts by finding the central axis. After extracting the breast ROI image, the horizontal distance is calculated. Then, the central axis is set as the half of the width of this horizontal distance. According to the central axis value, breast thermogram image is segmented as left and right images of breast. In case of asymmetric breast, as a women left and right breasts are not always symmetric, after extracting the initial breast ROI image, it will be converted into a binary image by using a threshold value to differentiate body from background. Then, full column with zero value, which represents the black Splitter line between right and left breasts, will be removed in order to divide the breast region into right and left. This can be done easily by using methods in image processing tool in MATLAB. Figure-6 shows the summary of the breast part boundaries extraction method.

2.4. Hough transform (HT) algorithm

In this section, breast ROI subjects to which called Hough Transform (HT) algorithm to accurately localizing only the right and left breast contours. The HT algorithm is used to identify the breast contour and also identifying the nipple circle part in the breast. The

generalized HT (GHT) is a powerful algorithm that is used to detect parametric patterns such as lines, circles and ellipses that presented in digital images. The concept of GHT is the transformation from image space (x and y coordinates) to a parameter space, representing the parameters of the desired shape. The parameter space is presented as m-dimensional accumulator array where m corresponds to the number of parameters considered. The authors in [32] proposed this concept in 1972 to identify and determine lines and curves in pictures, and Ballard [33] generalized the concept to detect an arbitrary shape.

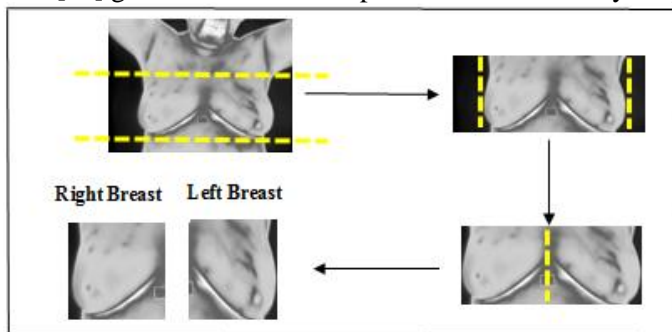


Fig. 6. The summary of the Breast Part Boundaries Extraction method.

In this paper, the HT is utilized to detect breast boundaries in thermograms which are like parabola shapes and also the circle shape of the nipples. So, the parabolic functions must be used to make the lower breast boundaries by using the two highest peaks coordinates in the parameter space of the HT adapted for the identification of parabolic curves. Also, the circle function should also be used to extract accurately the circle shape of the nipples. We found that the HT is very suitable for extracting curves and shapes that form the edges and limits of the breasts. The parabolic equation is represented as follows:

$$[(Y - Y_o)\cos\theta - (X - X_o)\sin\theta]^2 = 4p[(Y - Y_o)\sin\theta - (X - X_o)\cos\theta] \quad (3)$$

where,

Y_o, X_o are the parabola vertex coordinates.

θ is the angle of the parabola in polar coordinates.

p is the distance between vertex and focus of parabola.

Also, the circle equation is represented as follows:

$$(x - a)^2 + (y - b)^2 = r^2 \quad (4)$$

where,

(a, b) is the center of the circle.

r is the radius.

The determination of parabola breast curves and circle shape of the nipples using the GHT is performed by the following algorithm (2). Figure-7, shows both right and left breast contours.

Algorithm (2) Apply Hough Transform Algorithm.

1. Read the left breast and right breast images after applying canny edge detector.
 2. Enter the edge points into HT for parabola and circle detection.
 3. Convert all points in image space into curves in parameter space.
 4. Identify local maxima from the parameter (or accumulator) array. In other words, the local maxima in the accumulator array are compared with global maximum (GM) value and only whose values are equal to or greater than some fixed percentage of GM value are taken.
 5. De-Houghing into Cartesian space yields a collection of parabola descriptions and circle shapes of the image subject.
-

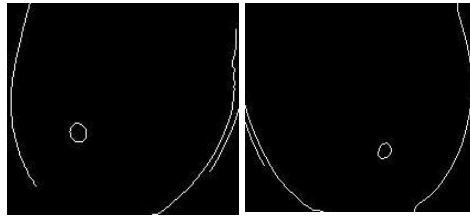


Fig. 7. Right and Left Breast Contours.

2.5. Boundaries contrast enhancement

In cases where edge boundaries are not extracted at all or small extra points need to be removed, the detected right and left breast contours must be enhanced. In this paper, the breast part image contrast is first improved by using a disk of 5-pixel radius to perform top-hat and bottom hat morphological filtering operations using MATLAB functions, thereby emphasizing the contours and the edges in the images. Also, the Canny edge detector threshold is determined adaptively by fixing a lower threshold on the contrast of the edges detected, i.e. the ratio of the pixels in edges to the total number of pixels in the region. Thus, the low-density breast edges have been improved by using the discussed method. For instance, Figure-8a shows an example of not accurate left and right breast contours. After applying the previous morphological filtering operations, the output image has been enhanced as found in [Figure-8b].



Fig. 8. (a) The low-density breast contours, (b) The enhanced density of breast contours.

3. Features extraction

The features extraction process is considered the most critical phase of the diagnosis of BC. After separating the breast part from the remainder of thermogram, some required features are extracted. The features are necessary to identify properties of an image region by utilizing space relationships related to the gray-level distribution of a given image. In this work, two types of features (first order statistical and texture) are extracted from the enhanced ROI of breast thermograms. The first order statistical features include mean,

variance, standard deviation, skewness and kurtosis. For texture features, the texture of an image region is calculated according to the gray levels distribution over the pixels in the region. However, the texture ones consist around twenty features which extracted from gray-level co-occurrence matrix (GLCM) [34] with a distance parameter of value ($d = 1$). These features are called Energy, Contrast, Correlation, Cluster Prominence, Cluster Shade, Entropy, Homogeneity, Dissimilarity, Maximum probability, Inverse difference moment (IDM), Inverse difference normalized (INN), Information measure of correlation1 (IMC1), Information measure of correlation2 (IMC2), Autocorrelation, Difference entropy, Difference variance, Sum of square variance, Sum entropy, Sum variance and Sum average. So, our total extracted features are 25 features.

3.1. First-order statistics

The first order texture measures are statistics evaluated from ROI breast image values, like mean, variance, etc., and do not consider relationships between pixels [35]. First order statistics are analyzed using the intensity histogram features. The histogram is a diagram representing the amount of pixels in an image at each different intensity value found in that image. Let random variable X represents the gray levels of breast image region. The first-order histogram $P(X)$ is defined as:

$$P(X) = \frac{\text{number of pixels with gray level } X}{\text{total number of pixels in the breast region}} \quad (5)$$

According to the definition of histogram $P(X)$, the required statistical features can be calculated. Mean is the first statistical features and can be defined as the average color in the image. It can be determined by the following formula:

$$\text{Mean: } m = \sum_{X=0}^{N_g-1} XP(X) \quad (6)$$

Where N_g is the number of possible gray levels? The variance defines the value of histogram width that calculates the deviation of gray levels from the mean. It is evaluated by using the following equation:

$$\text{Variance: } \mu_v = \sum_{X=0}^{N_g-1} (X - m)^2 P(X) \quad (7)$$

The Standard deviation (σ) is estimated by taking the square root of the variance in equation (6). Skewness helps to measure the degree of histogram asymmetry around the mean and equation (8) is used to calculate its value:

$$\text{Skewness: } \mu_s = \frac{1}{\sigma^3} \sum_{X=0}^{N_g-1} (X - m)^3 P(X) \quad (8)$$

Kurtosis formula represents a measure of the sharpness of histogram.

$$\text{Kurtosis: } \mu_k = \frac{1}{\sigma^4} \sum_{X=0}^{N_g-1} (X - m)^4 P(X) \quad (9)$$

All above equations are from [35]. Each breast region image is split into left and right parts. The first order statistical features are extracted to both of them.

3.2. Texture features

In this paper, the texture features are deduced from the Gray Level Co-occurrence Matrix (GLCM). The GLCM is a statistical method of investigating texture which takes into account the spatial relationship of pixels. It is a two-dimensional array which

considers the specific position of a pixel relative to the neighboring pixels [36]. The concept of this method depends on the determination of the second order statistical probability values represented by the matrix element $P(i, j | d, \theta)$ depending on the variations between gray levels 'i' and 'j' at a certain distance d and at a certain angle (θ). Each element (i, j) of GLCM represents the addition of the times that the pixel with value i placed in the specific spatial relationship to a pixel with value j in the input breast image.

There are about twenty features with parameters' values $d=1$ and (θ) are computed in four directions: horizontal, vertical, and two diagonal directions that are commonly used [37], [38], [39]:

Firstly, the following notations are used in various formulas of GLCM features:

$P_x(i)$ – i th input in the marginal-probability matrix acquired by adding the rows of $P(i, j)$.

$$P_x(i) = \sum_{j=0}^{N_g-1} P(i, j) \quad (10)$$

$$P_y(j) = \sum_{i=0}^{N_g-1} P(i, j) \quad (11)$$

$$P_{x+y}(k) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i, j) \cdot i + j = k. \quad k = 2.3.4. \dots .2N_g \quad (12)$$

$$P_{x-y}(k) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i, j) \cdot |i - j| = k. \quad k = 0.1. \dots .N_g - 1 \quad (13)$$

$$\mu_x = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} iP(i, j) = \sum_{i=0}^{N_g-1} iP_x(i) \quad (14)$$

$$\mu_y = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} jP(i, j) = \sum_{j=0}^{N_g-1} jP_y(j) \quad (15)$$

$$\mu = \sum_{i, j} iP(i, j) \quad (16)$$

$$\sigma_x = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i - \mu_x)^2 P(i, j) \quad (17)$$

$$\sigma_y = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (j - \mu_y)^2 P(i, j) \quad (18)$$

where μ_x , μ_y , σ_x and σ_y are the means and standard deviations of P_x and P_y . The extracted texture features are:

- Energy

It is called as uniformity of Angular Second Moment (ASM) which measures the homogeneousness of any image and can be determined from the normalized GLCM. This feature can detect disorders in textures.

$$f_1 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i, j)^2 \quad (19)$$

- Contrast

It measures the spatial frequency of an image. This feature helps to measure amount of local variations present in an image.

$$f_2 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i - j)^2 P(i, j) \quad (20)$$

- Correlation

It presents a measurement of gray tone linear dependencies between the pixels at the specified positions relative to each other in breast image.

$$f_3 = \frac{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i \times j) \times P(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y} \quad (21)$$

- Cluster Shade

It presents a value of the skewness and uniformity of the matrix. The image is asymmetric, if this feature is high.

$$f_4 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \{i + j - \mu_x - \mu_y\}^3 P(i,j) \quad (22)$$

- Cluster Prominence

It measures the asymmetry in an image. The image is less symmetric only if this feature value is high. But when its value is low, a peak in the GLCM matrix around the mean values is found.

$$f_5 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \{i + j - \mu_x - \mu_y\}^4 P(i,j) \quad (23)$$

- Entropy

It is a measurement of statistical randomness or intensity allocation of image texture. When all the probabilities $P(i, j)$ are equal, the entropy value is the highest. Its value will be smaller when the entries in $P(i, j)$ are unequal.

$$f_6 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i,j) \log(P(i,j)) \quad (24)$$

- Homogeneity

It is also called Inverse Difference Moment. The closeness the distribution of elements in the GLCM according to the diagonal can be measured using this feature. Its values range is [0 1]. The value 1 of homogeneity is for diagonal GLCM.

$$f_7 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{P(i,j)}{1+(i-j)^2} \quad (25)$$

- Dissimilarity

The dissimilarity measure is robustly related to high spatial frequencies in the texture of breast ROI image.

$$f_8 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} |i - j| P(i,j) \quad (26)$$

- Maximum Probability

It is a measurement of the powerful response of Co-occurrence matrix and its range is [0 1].

$$f_9 = \text{MAX}_{i,j} P(i,j) \quad (27)$$

The remainder of texture features is secondary and derived from those listed above.

- Inverse Difference Normalized (IDN)

$$f_{10} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{P(i,j)}{1+|i-j|/N_g} \quad (28)$$

- Inverse Difference Moment Normalized (INN)

$$f_{11} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{P(i,j)}{1+(i-j)^2/N_g^2} \quad (29)$$

- Information measure of correlation1 (IMC1)

$$f_{12} = \frac{HXY-HXY1}{\max\{HX, HY\}} \quad (30)$$

- Information measure of correlation2 (IMC2)

$$f_{13} = (1 - \exp[-2(HXY2 - HXY)])^{\frac{1}{2}} \quad (31)$$

Where,

$$HXY = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i,j) \log(P(i,j)) \quad (32)$$

$$HXY1 = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i,j) \log(P_x(i)P_y(i)) \quad (33)$$

$$HXY2 = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_x(i)P_y(i) \log(P_x(i)P_y(i)) \quad (34)$$

- Autocorrelation

It represents the repeating patterns like presence of periodic signal unobvious due to noise.

$$f_{14} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (ij)P(i,j) \quad (35)$$

- Difference entropy

$$f_{15} = \sum_{i=0}^{N_g-1} P_{x-y}(i) \log(P_{x-y}(i)) \quad (36)$$

- Sum entropy

This feature presents a measurement of the average of disorders or complexities in the ROI image

$$f_{16} = \sum_{i=0}^{2N_g-2} P_{x+y}(i) \log(P_{x+y}(i)) \quad (37)$$

- Difference variance

$$f_{17} = \sum_{i=0}^{N_g-1} i^2 P_{x-y}(i) \quad (38)$$

- Sum variance

It represents a value of average of heterogeneities in the separated ROI image.

$$f_{18} = \sum_{i=0}^{2N_g-2} (i - f_{15})^2 P_{x+y}(i) \quad (39)$$

- Sum of squares: variance

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

$$f_{19} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i - \mu)^2 P(i,j) \quad (40)$$

- Sum average

This feature evaluates the average of the gray level within breast ROI image.

$$f_{20} = \sum_{i=0}^{2N_g-2} iP_{x+y}(i) \quad (41)$$

3.3. Features evaluation

Now, before using these extracted features as inputs for the classifiers in the classification step, they were analyzed using student's t hypothesis testing the mean of two independent samples [40] to increase the performance of the proposed SM. In this test, the P value is calculated for each feature according to the two samples which are: normal and abnormal cases. Then, the features that have P value less than 0.05 were selected for the classification process. According to this test, as shown in Table 1, only the Mean feature are statistically insignificant where P greater than 0.05 while the Variance, Standard deviation, Skewness and Kurtosis are found to be statistically significant where P less than 0.05. Also, as seen in Table 2, Correlation, Entropy, Homogeneity, Dissimilarity, and Information Measure of Correlation1 (IMC1) are found insignificant features while the Energy, Contrast, Cluster Shade, Cluster Prominence, Maximum Probability, Inverse difference moment (IDM), Inverse difference normalized (INN), Information measure of correlation2 (IMC2), Autocorrelation, Difference entropy, Difference variance, Sum entropy, Sum variance, Sum of squares: variance, and Sum average were found highly significant features where P less than 0.05. Therefore, only 19 out of 25 features were used in the classification step.

Table 1.

First order statistical features analysis for normal (NC) and abnormal (AbC) cases.

Feature	Average of NC	Average of AbC	P_Value
Mean	148.876	138.235	0.061
Variance	1275.28	1245.01	0.028
Standard deviation	35.224	34.902	0.035
Kurtosis	4.494	3.1275	0
Skewness	-1.1854	-0.401	0

4. Classification

Classification is the final step in the BC detection method which consists of two operations: training and testing. Hence, the extracted features are used as inputs into the classification algorithms to analyze the breast thermal images. In this paper, to test the performance of the proposed automatic SM, different classification algorithms are applied and their results have been evaluated Such as SVM, and ANN.

4.1. Support vector machine (SVM):

SVM is a powerful classification algorithm which implements classification tasks by constructing hyperplanes in a multidimensional space that separates values of various class labels [41]. If a hyperplane has the longest distance to the closest training data point of any class, then a beneficial division can be attained; since a larger margin denotes the classifier has a lower generalization error. When classes overlap with other, a hyperplane is chosen for minimizing the errors of data points along or across the boundary line between classes; these points are indicated as the support points or support vectors.

Table 2.

Textural features analysis for normal (NC) and abnormal (AbC) cases.

Feature	Average of NC	Average of AbC	P_Value
Energy	0.00068	0.00065	0.0455
Contrast	2576.8	2613	0.0023
Correlation	0.0038	0.0049	0.0541
Cluster Shade	4555.22	6601.34	0.0076
Cluster Prominence	15700000	16490118	0
Entropy	9.6629	9.6667	0.551
Homogeneity	0.02384	0.02376	0.0901
Dissimilarity	41.52574	35.8311	0.2361
Maximum Probability	0.00009	0.0001	0.0078
IDN	0.8829	0.9816	0.0263
INN	0.77659	0.87542	0.0241
IMC1	-0.00334	-0.00425	0.0982
IMC2	0.174203	0.19804	0.0234
Autocorrelation	4185.636	4243.11	0.0008
Difference entropy	4.636513	5.642285	0.0081
Sum entropy	5.316307	5.328469	0.0316
Difference variance	2576.802	2612.99	0.0037
Sum variance	17971.01	18226.3	0
Sum of squares: variance	5516.292	6556.176	0.0014
Sum average	129.4688	130.1985	0.0237

4.2. Artificial neural network (ANN):

ANN is considered as a powerful classification tool that can represent complex input-output relationships. ANN can also represent as a Multilayer Perceptron (MLP) with Backpropagation Network [42] which consists of at least three layers: single input layer, at least one hidden layer, and single output layer. An input pattern is transmitted in forward direction to output units when a cycle of backpropagation takes place. As the method of supervised learning, the errors in weights and biases transmit backwards from output nodes to the inner nodes.

5. Experimental results

5.1. BC dataset

A benchmark breast database is selected from an open online data base PROENG (<http://visual.ic.uff.br/>) which called DMR-IR database [27]. It is used to estimate the proposed approach. This paper used a 200 (90 normal and 110 abnormal) IR breast images (640×480 pixels) from this database. As known, the proposed method must be tested on the database. Therefore, total of 130 samples of thermograms are used to train the network and remaining 70 samples of thermograms are used to test and validate the proposed SM.

5.2. Experimental results

In this section, two types of results are presented: the segmentation results which obtained due to the new automatic SM and the classification results based on this new method.

5.2.1. Segmentation Results:

- Qualitative Segmentation Results:

According to the different breast sizes between women, the proposed SM is applied to various cases of thermograms. The segmentation results of different cases of breast thermograms (i.e., different breast sizes) using our new method are introduced in [Figure-9]. It appears the automatic separated breast ROI image from thermograms for the various breast sizes.

- Quantitative Segmentation Results:

The task of segmentation evaluation is done by comparing two segmentations by measuring the similarity between them as one is the segmentation to be evaluated and the other is the corresponding groundtruth segmentation. The performance of the proposed SM is validated with breast overlapping metrics such as Dice Index (DI), the Jaccard index (JI), false positive (FP) and false negative (FN) errors. The DI [43] [44] is considered as the most used metric in validating medical segmentation results. It provides a direct comparison between automatic and ground truth segmentations. It can be calculated by using the following equation:

$$DI(G.S) = 2 \frac{|G \cap S|}{|G| + |S|} \quad (42)$$

where S and G denote the segmentation result and the corresponding ground truth, respectively. The JI [43] [45] is another important metric which is defined as the intersection between them divided by their union. It can be determined by the following equation:

$$JI(G.S) = \frac{|G \cap S|}{|G \cup S|} \quad (43)$$

The FP and FN represents the number of false positives and the number of false negatives, respectively. They can be calculated as follows:

$$FP = \frac{|S| - |G \cap S|}{|G|} \quad (44)$$

$$FN = \frac{|G| - |G \cap S|}{|G|} \quad (45)$$

These equations give values that range from 0 to 1. The higher values indicate better performance only for DI and JI , while a value of zero is achieved when performing perfect breast volume segmentation for FP and FN. Table 3 shows the quantitative segmentation results for normal and abnormal cases. These results show that the proposed SM achieves a high segmentation accuracy.

Table 3.

The quantitative segmentation results for normal and abnormal cases.

Cases	DI	JI	FP	FN
Normal	0.9762	0.945	0.007	0.005
Abnormal	0.9657	0.951	0.012	0.0029

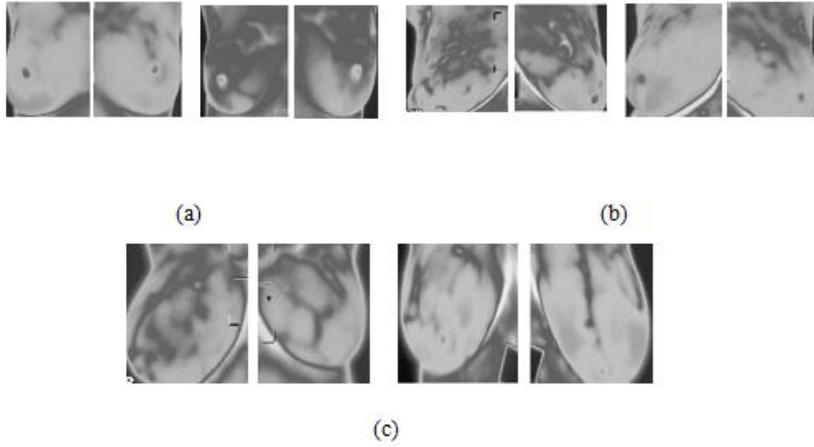


Fig. 9. (a) Small breast sizes, (b) Big breast sizes, (c) Asymmetric breast sizes.

5.2.2. Classification results

In this work, MLP-based and SVM-based approaches are applied to our BC dataset for diagnosis of BC. The class attribute has got only two values, namely Normal and Abnormal. All the selected features that form data set along with their values range are reported in Table1 and Table2. MLP and SVM classifiers were applied on the BC data set for simulation by the WEKA software. WEKA software helps to create the BC detection model.

The simulation results are introduced in the Table 4. Table 4 mainly summarizes the results founded on different performance measures such as Classification Accuracy, Root mean squared error (RMSE), Kappa statistic, True Positive Rate (TP-Rate), False Positive Rate (FP-Rate), Precision, Recall, F-Measure, Dice Index (DI), the Jaccard Index (JI), and the Jaccard Distance Index (JDI). These several standard terms have been defined for the 2 class confusion matrix (Normal and Abnormal):

The Accuracy is important to determine the overall correctness of the model and it is calculated by dividing the sum of correct classifications over the total number of classifications, as found in the following equation:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (46)$$

The Precision is the rate of the positive cases that were predicted correctly, and is determined by the following equation:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (47)$$

The FP-Rate is the proportion of negative cases that were classified as positive incorrectly and can be determined by the equation:

$$\text{FP - Rate} = \frac{FP}{FP+TN} \quad (48)$$

The Recall or TP-Rate is the proportion of the correctly identified positive cases, and is determined using the equation:

$$\text{Recall} = \text{TP - Rate} = \frac{TP}{TP+FN} \quad (49)$$

The higher precision may be very important for many cases, and in other cases higher recall may be very important. However, in most cases, we aim to improve both values. F-Measure is said to be the collection of these values, and determined as in following equation:

$$F - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (50)$$

The Dice Index (DI) [44] is an important metric for validating the performance of classification. It measures the similarity of two regions and ranges from 0 to 1. It is calculated by

$$DI = \frac{2TP}{2TP + FP + FN} \quad (51)$$

Another important metric that has been considered for evaluation of classification results is the Jaccard Index (JI) [45], which describes the relation between correct detection and failures. The result becomes better if the JI is closer to one. It is represented here as follows:

$$JI = \frac{TP}{TP + FP + FN} \quad (52)$$

In contrast to JI, the Jaccard Distance Index (JDI) is a measure of dissimilarity. It is calculated by subtracting JI from one ($1 - JI$), such that the best result is occurred when the JDI value closes to zero.

Table 4.

The various performance measures results of SVM and MLP classifiers.

Classifier	Accuracy	RMSE	Kappa statistics	TP-Rate	FP-Rate	Precision	Recall	FMeasure	DI	JI	JDI	Class
SVM	96.667%	0.2314	0.9331	0.984	0.053	0.954	0.984	0.969	0.969	0.939	0.061	Normal
				0.947	0.016	0.982	0.947	0.964	0.964	0.939	0.061	Abnormal
MLP	96.071%	0.1982	0.9213	0.934	0.008	0.993	0.934	0.963	0.963	0.928	0.072	Normal
				0.992	0.066	0.927	0.992	0.958	0.958	0.928	0.072	Abnormal

6. Discussion

According to the segmentation results, Figure-10 shows that our proposed automatic SM can successfully separate the ROI breast image from thermograms with various breast sizes. The features that extracted from the ROI images have been classified into normal and abnormal. The classification step results after using SVM and MLP classifiers are summarized in Table3. After using our proposed SM, the accuracy of SVM and MLP reached 96.667%, and 96.071%, respectively. Also, all other values of performance measures in Table3 shows that our BC detection models (using SVM and MLP classifiers) are more effective for the analysis of breast thermograms. For example, using the measurements of the recall and the precision, the classification results reached to about 100%. This obtained highest accuracy denotes that the proposed automatic SM is a favorable technique for extracting the breast ROI image from breast thermograms.

7. Conclusions

This paper presents a new automatic SM for the BC thermal images. This SM is used to separate the breast region out from the captured original breast thermograms and divide this region into left breast and right breast ROI images. To do that, four limit lines (upper, lower, left, and right) were identified automatically and breast region was extracted from

thermogram based on these lines. After that, breast thermogram image was segmented as left and right breast image depending on the central axis value. Then, the HT algorithm was used for accurate identifying and enhancing the breast boundaries in the output ROI image. The results of the proposed SM prove its reliability in extracting the breast ROI from original thermal images for different cases. After segmentation, statistical significant features were extracted from the separated breast ROI image. Finally, to evaluate the output results of this new method, the features that extracted from breast ROI images gave as entries to SVM and MLP classifiers to detect the normal and abnormal breasts. The experimental results showed that the accuracy of SVM and MLP classifiers reached to 96.667% and 96.071%, respectively. Also, using the measurements of the recall and the precision, the classification results reached to about 100%. These excellent classification results, based on our new SM, prove that our SM could be used for the automatic segmentation of the thermal BC images. In future work, to enhance the performance of the proposed SM, it should handle also the case of slanted breast thermograms. Also, the effect of using as example a Particle swarm optimization technique as a feature selection search method can be evaluated. This method can help to test the effect of the reduction of the whole extracted features into smaller feature subset on the classifier performance.

REFERENCES

- [1] L. A. Torree, "Global cancer statistics, 2012," *CA: a cancer journal for clinicians*, Vol. 65, No. 2, pp. 87-108, 2015.
- [2] INCA. Instituto Nacional de Câncer. <http://www2.inca.gov.br>, 2012.
- [3] Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2016. Toronto, ON: Canadian Cancer Society; 2016. www.cancer.ca/statistics.
- [4] American Cancer Society, "Cancer Facts & Figures 2017", American Cancer Society Inc., pp. 1-71, 2017. www.cancer.org.
- [5] World Health Organization: WHO 2014b. Available from: URL: <http://www.who.int>.
- [6] E. Y. K. Ng, and N. M. Sudharsan, "Numerical computation as a tool to aid thermographic interpretation," *Journal. Of. Medical. Eng. and Technology*, Vol. 25, pp. 53-60, 2001.
- [7] Aya Hossam, Hany M. Harb, and H. M. Abd El Kader, "Performance Analysis of Breast Cancer Imaging Techniques," *International Journal of Com. Science and Inf. Security*, Vol. 15, No. 5, pp. 48-56, May 2017.
- [8] M. R. K. Mookiah, U. Rajendra Acharya, and E.Y.K. Ng, "Data mining technique for Breast Cancer detection in thermogram using hybrid feature extraction strategy," in *Qualitative InfraRed Thermography*, Vol. 9, No. 2, pp. 151-165, December 2012.
- [9] V. Umadevi, S.V. Raghavan, and S. Jaipurkar, "Framework for estimating tumor parameters using thermal imaging," *Ind. Journal of Medical Research.*, Vol. 134, No. 5, pp. 725-731, 2011.
- [10] M. Y. Lee, and C. S. Yang, "Entropy Based Feature Extraction and Decision Tree Induction for Breast Cancer Diagnosis with Standardized Thermograph Images," *Comp. Methods and Prog.in Biomedicine*, Vol. 100, no. 3, pp. 269-282, 2010.
- [11] M. Løberg, M. L. Lousda, M. Bretthauer, and M. Kalager, "Benefits and harms of mammography screening," *Løberg et al. Breast Cancer Research*, Vol. 17, No. 63, 2015.
- [12] M. G. Marmot, D. G. Altman, D. A. Cameron, J. A. Dewar, S. G. Thompson, M. Wilcox, and The Independent UK Panel on BC Screening, "The benefits and harms of BC screening: an independent review", *British Journal of Cancer*, Vol. 108, No. 11, pp. 2205-2240, 2012.
- [13] D. Walker, and T. Kaczor, "Breast thermography: History, theory, and use," *Natural Medicine Journal*, Vol. 4, No. 7, Jul. 2012.
- [14] D. Kennedy, T. Lee, and D. Seely, "A Comparative Review of Thermography as a Breast Screening Technique," *Integrative Cancer Therapies*, Vol. 8, No. 1, pp. 09-16, 2009.

- [15] E.Y.K. Ng, "A review of thermography as promising non-invasive detection modality for breast tumor," *Int. Journal of Thermal Sciences*, Vol. 48, no. 5, pp. 849-859, 2009.
- [16] P. Kapoor, S.V.A.V. Prasad and S. Patni, "Automatic Analysis of Breast Thermograms for tumor detection based on Biostatistical feature extraction and ANN," *IJETED*, Vol. 7, no. 2, pp. 2249-6149, 2012.
- [17] V. R. Udupi, S. S. Tamboli, V. A. Patil, and A. S. Tamboli, "Histopathological Image Analysis for BC Dignosis: A Review," *IJECT*, Vol 2, no. 4, pp.164-170, 2011.
- [18] C. Lipari, and J. Head, "Advanced infrared image processing for BC risk assessment," *Proc 19th Int Conf IEEE/EMBS, Chicago IL*, pp. 673–676,1997.
- [19] S. Suganthi and S. Ramakrishnan, "Semi-automatic segmentation of breast thermograms using variational level set method," in the 15th International Conference on Biomedical Engineering. Springer, Vol. 43, pp. 231–234, 2014.
- [20] S. S. Srinivasan and R. Swaminathan, "Segmentation of breast tissues in infrared images using modified phase based level sets," in *Biomedical Informatics and Technology*. Springer, pp. 161–174, 2014.
- [21] Q. Zhou, Z. Li, and J. K. Aggarwal, "Boundary extraction in thermal images by edge map," in *Proc. of the 2004 ACM sym. on Applied computing*, pp. 254–258, 2004.
- [22] M. E. Tavakol, V. Chandran, E. Y.-K. Ng, R. Kafieh, "BC detection from thermal images using bispectral invariant features," *Inter. Journal of Thermal Sciences*, 69, pp. 21-36, 2013.
- [23] P. Soura, B. Debotosh, and N. Mita, " Wavelet Based Thermogram Analysis for BC Detection," *IEEE Int. Sym. on Advanced Computing and Communication*, pp. 205-212, 2015.
- [24] U. Acharya, E. Y.-K. Ng, J.-H. Tan, and S. V. Sree, "Thermography based BC detection using texture features and support vector machine," *Journal of med. Sys.*, Vol. 36, No. 3, pp. 1503–1510, 2012.
- [25] M. C. Araújo, R. C. Lima, and R. -M. De Souza, "Interval symbolic feature extraction for thermography BC detection," *Expert Sys. with -App.*, Vol. 41, No. 15, pp. 6728–6737, 2014.
- [26] S. V. Francis, M. -Sasikala, and S. Saranya, "Detection of breast abnormality from thermograms using curvelet transform based feature extraction," *Journal of med. Sys.*, Vol. 38, No. 4, pp. 1–9, 2014.
- [27] PROENG (2012). Image processing and image analyses applied to mastology. <http://visual.ic.uff.br/en/proeng/>.
- [28] Deserno TM. -Biological and -Medical Physics, -Biomedical -Engineering Greenbaum E, editor.: Springer-Verlag Berlin Heidelberg; 2011.
- [29] Canny, and John, "A Computational -Approach to -Edge -Detection," *IEEE Trans. on -Pattern -Analysis and -Machine -Intelligence*, Vol. PAMI-8, No. 6, pp. 679-698, 1986.
- [30] P. P. Acharjya, R. Das, and D. Ghoshal, "Study and -Comparison of Different Edge Detectors for Image Segmentation," -*Global -Journal of -Computer -Science and Technology*, Vol. 12, No. 13, pp.29-32, 2012.
- [31] L. Silva, D. -Saade, G. -Sequeiros, A. Silva, A. -Paiva, R. -Bravo, and A. Conci, "A new database for breast research with infrared image," *Journal of -Medical Imaging and -Health -Informatics*, Vol. 4, No. 1, pp. 92–100, 2014.
- [32] R. O. Duda, and P.E. Hart, "Use of the Hough transformation to detect lines and curves in pictures", *Commun ACM.*, Vol. 15, No. 1, pp. 11–15, 1972.
- [33] D. H. Ballard, "Generalizing the Hough -transform to detect arbitrary shapes.", *Pattern Recognit*, Vol. 13, No. 2, pp. 111–122, 1981.
- [34] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, Vol. SMC-3, No. 6, pp. 610–621, 1973.
- [35] A. Gupta, "Groundwork of Mathematical Probability and Statistics, " 3 td Edition, Academic Publishers, Calcutta, India (1995).

- [36] A. MohdKhuzi, R. Besar, W. Zaki, and NN. Ahmad, "Identification of masses in digital mammogram using gray level co-occurrences matrices," *Biomedical Imaging and Intervention Journal*, Vol. 5, No. 3, pp. 1-13, 2009.
- [37] A. Baraldi, and F. Parmiggiani, "An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters", *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 33, No. 2, pp. 293304, 1995.
- [38] M. Partio, B. Cramariuc, M. Gabboui, and A. Visa, "Rock Texture Retrieval using Gray Level Co-occurrence Matrix", *Proceedings of 5th Nordic Signal Processing Symposium*, 2002.
- [39] M. Tuceryan, and A. K. Jain, "Texture Analysis, The Handbook of Pattern Rec. and Computer Vision, Ed. 2", *World Scientific Publishing Co.*, 1998.
- [40] R. Tello, and P. E. Crewson, "Hypothesis Testing II: Means", *Radiology*, pp. 1-4, April 2003.
- [41] C. Cortes, and V. Vapnik, "Support-vector networks," *Machine learning*, Vol. 20, No. 3, pp. 273–297, 1995.
- [42] R. Rojas, "Neural Networks: A Systematic Introduction," *Springer*, 1996.
- [43] A.A Taha, A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imaging*, Vol. 15, No. 29, pp.1-28, 2015
- [44] L.R. Dice, "Measures of the amount of ecologic association between species", *Ecology*, Vol. 26, No. 3, pp. 297–302, 1945.
- [45] P. Jaccard, "The distribution of the flora in the alpine zone", *New Phytologist*. Vol. 11, No.2, pp.37–50, 1912.

"طريقة لتقسيم الصورة تلقائيا لتحليل سرطان الثدي باستخدام التصوير الحراري" الملخص العربي:

يعد سرطان الثدي واحد من أكثر الأمراض شيوعا بين النساء في جميع أنحاء العالم مما يؤدي إلى عدد كبير من وفيات. و يعد الكشف المبكر عن هذا المرض ضروري لتعزيز فرصة البقاء على قيد الحياة. و من هم الأدوات الفعالة التي تساعد في الكشف المبكر عن سرطان الثدي هي التصوير الحراري. و لإجراء التصوير الحراري للثدي يتم استخدام كاميرا الأشعة تحت الحمراء الحرارية التي تحول الأشعة تحت الحمراء المنبعثة من جسم الإنسان إلى إشارات كهربائية، وتقديمها على شكل صورة حرارية للثدي. ولتحقيق نظام فعال للكشف عن سرطان الثدي باستخدام التصوير الحراري، فمن الضروري استخراج المنطقة ذات الاهتمام المحتمل إصابتها في الصور الحرارية وهي منطقة الثدي فقط قبل تحليلها. لذلك تقدم هذه المقالة طريقة تجزئة أوتوماتيكية جديدة لإستخراج الصورة الخاصة بالمنطقة المهمة فقط من الصورة الحرارية للثدي لتحليلها. و بعد استخراج هذه الصورة باستخدام طريقة التجزئة المقترحة يتم استخدام طريقتين مختلفتين للتحديد الدقيق لحدود الثدي في الصورة المستخرجة و هما خوارزمية تحويل هوف و التي تقوم بتأكيد تحديد منحنيات الثدي في الصورة المستخرجة و الطريقة الأخرى تقوم بتحسين التباين و زيادة كثافة حدود الثدي في الصورة المستخرجة. و لتقييم نتائج هذه الطريقة المقترحة، تم استخراج الخصائص الإحصائية من صورة الثدي المستخرجة من الصورة الحرارية. و بعد ذلك تم استخدام المصنفات مثل أداة دعم القوة و الشبكة العصبية الاصطناعية للكشف عن سرطان الثدي و تحديد الحالات الطبيعية والغير طبيعية اعتمادا على الخصائص الإحصائية المستخرجة. و أثبتت النتائج التجريبية أن هذه المصنفات المستخدمة حققت نسبة دقة عالية تصل إلى 96.67% في حالة أداة دعم القوة و 96.07% في حالة الشبكة العصبية الاصطناعية. و هذه النتائج تشير إلى أن الطريقة المقترحة للتجزئة التلقائية للصور الحرارية للثدي هي طريقة فعالة و يمكن الإعتماد عليها للحصول على دقة عالية في الكشف عن سرطان الثدي.