Ain Shams Engineering Journal xxx (xxxx) xxx

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Engineering Physics and Mathematics

Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques

Mai S. Mabrouk [a,*], Heba M. Afify [b], Samir Y. Marzouk [c]

[a] Department of Biomedical Engineering, MUST University, Egypt
[b] Department of Bioelectronics Engineering, MTI University, Egypt
[c] Departement of Basic and Applied Science, Arab Academy of Science and Technology, Egypt

ARTICLE INFO

Article history:
Received 29 July 2018
Revised 12 November 2018
Accepted 22 January 2019
Available online xxxx

Keywords:
Computer-aided diagnosis (CAD)
Mammograms
K-nearest neighbor classifier (KNN)
Support vector machine (SVM)
Artificial neural network (ANN)

ABSTRACT

Computer-aided diagnosis (CAD) approach is presented as strong frameworks to solve the inaccuracy problems. The major purpose of this paper is to improve a CAD system depended on supervised classification that can be useful in diagnosing and detecting the changes of breast cancers in digitized mammograms earlier, accurately and faster than standard examination programs by applying CAD according to image processing techniques beginning with preprocessing step, segmentation, feature extraction and finally classification stage. The work presented in this study is based on the integration of different features such as shape, texture and invariant moment features. This integration achieved best results for sensitivity and specificity rather than using the one type of features in breast cancer classification. The accuracy of our integration system reached 96% in the automatic mode of ANN while best accuracy accomplished by features result according to invariant moments that reached 97% by ANN in an automatic way.

1. Introduction

Generally, breast cancer is spread due to DNA damage or hereditary changes that threatened the women health [1,2]. The important factors that control the appearing breast cancer are gender, age, and genetic factor. Mammography is imaging instrument of the breast to distinguish between benign and malignant.

Breast cancer has several special diseases like microcalcifications (MCCs), which are very small bits of calcium, appearing on a mammogram as small bright spots and masses. MCCs are tiny sizes from 0.33 to 0.7 mm and are shinier than neighboring tissues. Therefore, MCCs are hard to discover by the radiologist in examination programs because they show with low contrast due to their little shape [3]. The radiologist decided the cancer presence or not according to the shape of MCCs. However, there are 30–50% errors in breast cancer diagnostics early because of the existence of MCCs clusters.

Early detection and recognition of breast cancer played a significant function in cancer handling to reduce the death-rate [4]. An interesting solution to getting more information from the breast, and helping the physicians to raise their detection accuracy is using CAD [5]. CAD algorithms have been used for image analysis to form the feature extraction in order to achieve the correct decisions [6]. The computer outcome is more accurate than radiologist's diagnosis because of differences in diagnosis can be large [7]. Several methods in mammography have been used to promote radiologists diagnostic and detection performance by indicating suspicious areas. Stefanoyiannis et al. [8] suggested a model for mammographic image techniques. This technique involved equalization of images density by improving the contrast at the breast periphery.

Byng et al. [9] presented image enhancement by using filtering technique to clarify the breast images. Pohlman et al. [10] prospered a high level of detection sensitivity for breast images by an adaptive region growing technique. Petrick et al. [11] suggested two stages of image enhancement by Difference-of-Gaussian (DoG) edge detector for the masses discovery. Kegelmeyer et al.

* Corresponding author.
E-mail address: Msm_eng@yahoo.com (M.S. Mabrouk).

Peer review under responsibility of Ain Shams University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.asej.2019.01.009
2090-4479/

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009

Nomenclature

CAD	computer-aided diagnosis
MCCs	micro calcification clusters
KNN	K-nearest neighbor classifier
SVM	support vector machine classifier
ANN	artificial neural network classifier
DoG	difference-of-gaussian
ROI	region of interest
MIAS	mammographic image analysis society
FSHS	full-scale histogram stretching
HE	histogram equalization techniques
LT	local thresholding
MLO	mediolateral oblique view
GLCM	gray level co-occurrence matrix
LEOH	local edge orientation histogram
RBF	radial basis function
RBFNN	radial basis function neural network

[12] improved a technique for speculated masses detection of 5 features sets for each pixel by local edge orientation histogram (LEOH) and the output of four spatial filters which are a subset of k. Laws texture features [13].

The objective of this paper is to build a CAD system that recognizes and describes techniques of MCCs detection. The proposed methodology is based on several steps; a CAD system started through preprocessing, segmentation, feature extraction, and finally classification.

2. Material and methods

The most common databases are used in mammography research areas that contain data warehouse from the mammographic image analysis society (MIAS) [14,15]. MIAS database includes 181 breast images for 161 patients, which divided into three types classified as 97 normal, and 84 benign/malignant.

In this paper, the CAD system detected mass lesions and MCCs by analyzing the breast images through two steps. Detection of suspicious area is the first step and the second step has segmented this area into parts and features are evaluated for each part. These features also used to define whether the lesion is benign or malignant. Initially, there are some popular steps that have been used to define the suspicious lesions. Fig. 1 displayed an exemplary plan for CAD system that starting from the preprocessing step of the mammogram database. Then, the breast region is segmented and image processing techniques may utilize to improve the quality of the image and minimize the noise. The region of interest (ROI) acts as a selection step, where a group of dubious ROIs is chosen to identify them as usual or unusual. After that, a feature extraction step is calculated for the chosen ROI, where a set of features is studied for the extracted ROI. Hence, feature selection is computed to serve in the classification process. It means that successful of classification process based on the selected features that have a strong influence on the model. Ultimately, a classification step is implemented in order to differentiate between normal and abnormal images.

2.1. Implementation

In this paper, all presented algorithms are implemented in MATLAB version 9.2 under name of R2017a and performed on a PC equipped with a 1.6 GHz-Intel Pentium processor and 512 MB of RAM. In MATLAB, images are represented as two-dimensional matrices, allowing low-level image analysis and manipulation using MATLAB's efficient matrix and vector computations. We used MATLAB to implement an automated image analysis system that takes advantage of the functions and features that described.

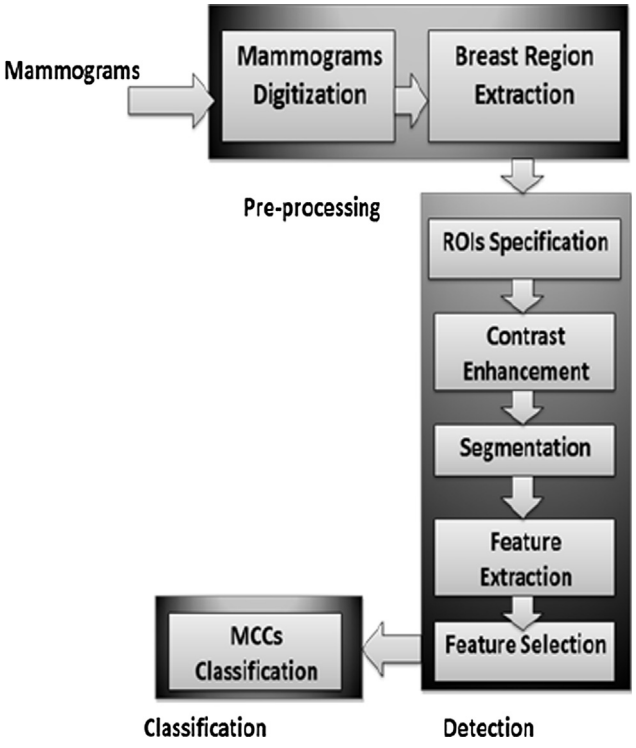

Fig. 1. A proposed framework for CAD system.

2.2. Image preprocessing

Preprocessing is a significant step in low-rate image treatment. The fundamental idea of preprocessing is to enhance the fineness of the original mammogram image, to grow the intensity difference between objects and background, and to make the following steps in CAD easier and more reliable, in order to manufacture trustworthy models of breast tissue frames. Although MCCs are usually shinier than their environments, their variance in a dense breast is extremely low so that human eyes can hardly recognize them. Therefore, the purpose of the preprocessing phase is to enlarge the variance of MCCs over a threshold to separate them from their environments.

2.3. Image enhancement

Image enhancement included gray-level manipulation, noise decreasing, contrast increasing, shape monitoring, and edge crisping by filtering, extraction, and enlargement. Image enhancement is applied to avoid low contrast of images and determine masses

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009

in images easily to help radiologists in the detection of abnormalities. This technique used in this paper for handling mammogram images as shown in Fig. 2, which showed four main categories applied to the original image, namely;

(1) Full-Scale Histogram Stretching (FSHS) which used to present the clear advancement in the optical quality of the image that suffering from a limited distribution of gray levels.
(2) Histogram Equalization techniques (HE) which used to improve low-contrast of the input image.
(3) Morphological Enhancement techniques which used to delete the noise from image to get the well-enhanced image.
(4) Wavelet Transform which used to implement different frequencies of a signal by different scales of the image to enhance the image quality and create the outcomes of segmentation more clear.

2.3.1. Full-Scale histogram stretching (FSHS)

Contrast stretching also called FSHS, is an easy linear point process that extends the image histogram to fill the whole available grayscale area [16]. FSHS was applied in [17] to improve mammogram contrast and visibility, allowing for more accurate contours and feature detection.

2.3.2. Histogram Equalization (HE)

HE is used for reducing the impact of over-brightness or over darkness to improve the optical look of images [18]. The HE technique is operative in improving the whole image with low contrast, especially if that image contains a single object or there is no evidence of contrast variance between the object and background. This is attained by applying a normalized accumulative histogram as a grayscale charting function. The basic function of HE technique is to reset the intensity degrees of pixels to gain a regular allocation of intensities over the full grayscale area.

2.3.3. Morphological enhancement

Essentially, morphology cares about the contour and structure of the image for image extraction, and noise elimination [19]. The idea of the morphological process is based on adding or removing pixels of the image boundaries. In the morphological enhancement, there are two processes for determining the number of pixels in the output image by studying the relations between corresponding pixel and its neighbors in the input image. We applied morphological to eliminate the bright background caused by dense breast tissues by maintaining the features of the masses. Therefore, background correction is an important phase for improvement and more accurate mass segmentation result.

We proposed dual morphological top-hat operations [20] which applied for noise deletion in the background and inside the suspicious mass to get the best results in image enhancement technique.

2.3.4. Wavelet transform

Filtration of the image is implemented by using low pass and high pass filters followed by a decimation operation [21]. By using

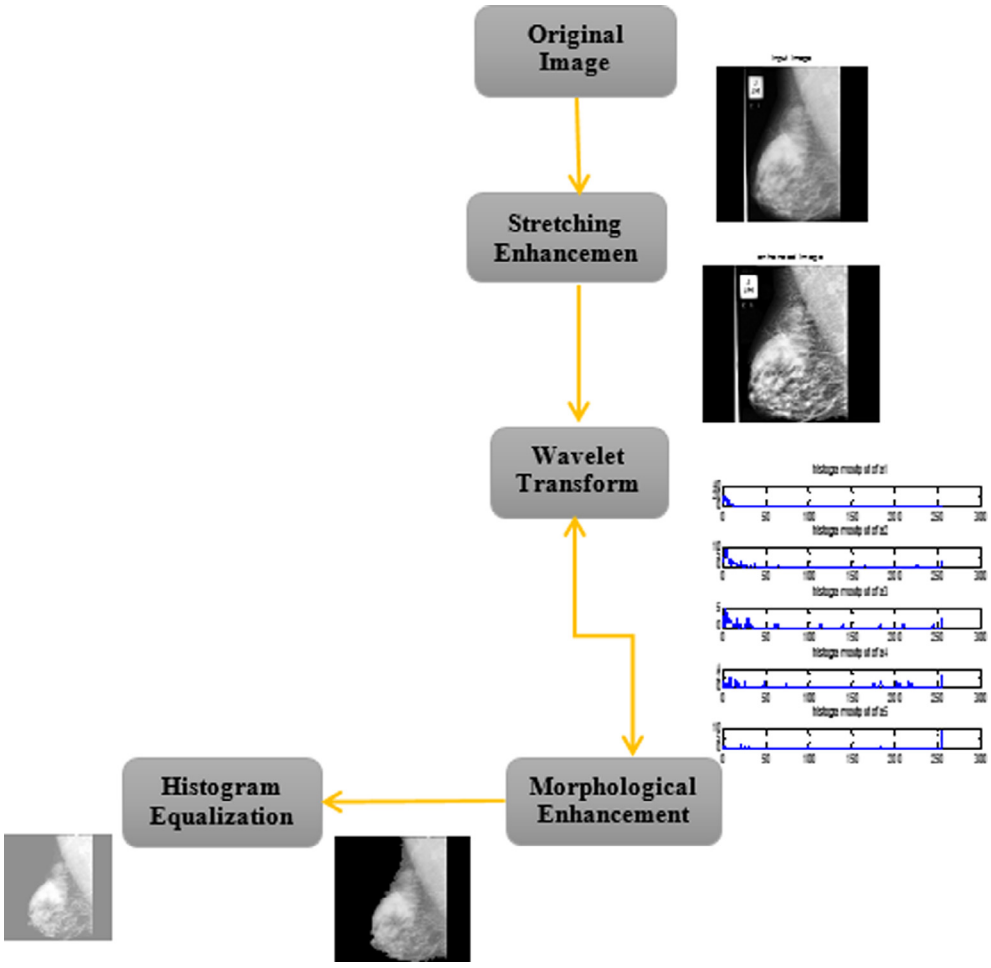

Fig. 2. Block diagram of breast image enhancement.

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009

this technique, the image is divided into two bands at each decomposition stage.

Because of the nature of mammogram, few levels of decompositions are necessary to analyze them [22]. In this paper, three-level decomposition and the Daubechies (6db) wavelet transform are used.

On the other hand, the wavelet technique has confirmed to be suitable in denoising. The proposed technique is implemented to vertical, horizontal, and diagonal detail components then computed the multi-scale threshold coefficients. The images in the scale-2 are issued for segmentation to reveal the current tumors in the digital mammograms. Next 5-scale wavelet transforms for the image histogram in scale-2 is used.

2.4. Image segmentation

The purpose of segmentation is to change the modeling of the image into something that is more expressive and simpler to analyze for definition every part such as lines and curves in the image. This is an essential step to extract the sensitivity of the whole image. The output of segmentation is included the areas containing all masses by collecting the segment sets or contour sets in the image [23]. For the segmentation step, we applied Local Thresholding (LT) Technique [24] and Otsu method [25] on breast images. LT is based on locating a single value known as the intensity threshold to exploit it for comparing each pixel with this value. It found that white area in the image indicates high pixel intensity rather than the threshold value. Also, we implemented a binarization process on the equalized image by Otsu method.

2.4.1. Pectoral muscle separation

Pectoral muscle separation is a critical key in mediolateral oblique view (MLO) for clarifying breast tissue. Extracting the pectoral muscle [26] is serious in mammogram image estimation by using the MATLAB imclearborder function. The pectoral muscle segmentation is a complex phase that is controlled by many factors such as shape and texture of muscle edge [27].

2.4.2. Background removal

For elimination of the background information like labels and wedges in the mammogram images, we first convert the grayscale or color image to binary image by using threshold technique and morphological operations. Secondly, remove the noise or background information from the binary image by removing all the smaller objects except the largest mammography part.

2.4.3. Fill in tumor

A combination of different morphology functions is used to overcome noise effects, which may disrupt the detection process of breast boundary. In the threshold process, not all lesions are clear-cut. Filling in spaces that found at lesion object completed tumor holes. ROI is obtained after filling holes of the resulting image.

2.4.4. Trace boundary

For extraction of the lesion contour, a simple blob-finding algorithm [28] is applied to the binary image previously gained. The boundary outline detected by collecting the edges of the black and white image in an arrangement table among the tracing technique. Tracing boundary is explained that nonzero pixels related to the object while zero pixels related to the background.

2.4.5. MCC extraction

Finally, the original image is multiplied by the last image resulting from the bounding box. It means that the border is composed of the image for formation of feature extraction.

An overview of the whole preprocessing steps is summarized through image representation of each step as shown in Fig. 3.

2.5. Feature extraction

The feature extraction stage has a high efficiency in CAD quality. Feature extraction proposed to reduce the original image by computing parameters for differentiating between ROI characteristics. We organized into morphologic features, texture features, and moment invariants features.

2.5.1. Morphological features

Morphological features are human notable and direct characteristics defined by radiologists. They mainly deal with the shape. As the MCC morphology in terms of physical shape is considered the major indication for the presence of ROI, it was valuable for the

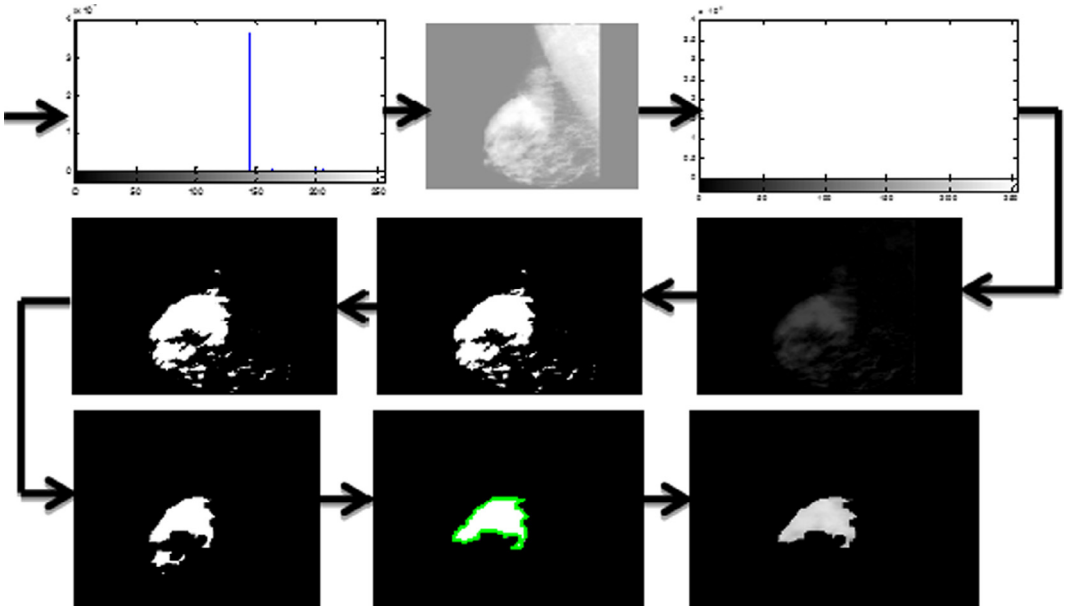

Fig. 3. Schematic representation of image preprocessing steps.

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009

proposed diagnostic tool to allow the recognition of morphologic structures that may not be visible. We selected shape features based on margin and density features plus 20 geometric features as described below:

1- *Area:*

Area measurement is considered as one of the easiest feature measurements, where it just requires counting the number of pixels occupied by the lesion. This gives an accurate measure of the area, as long the lesion is correctly segmented so that the white pixels just represent the required region.

2- *Perimeter:*

It is measured as the length of a region boundary.

3- *Solidity:*

It is a number of pixels by computing the total region area divided by the convex area.

4- *Circularity:*

It is computed by Eq. (1), where A and P represents area and perimeter, respectively.

$$Circularity = \frac{4\pi A}{P^2} \tag{1}$$

5- *Irregularity:*

There is a large number of ways to measure edge irregularities; some of them are more appropriate than others. One of the easiest ways of estimating edge irregularity is by using a measure called the compactness index as in Eq. (2).

$$Compactnessindex = \frac{P^2}{4\pi A} \tag{2}$$

6- *Major axis:*

It explains the major axis length of the labeled region of the lesion's binary image.

7- *Minor axis:*

It explains the minor axis length of the labeled region of the lesion's binary image.

8- *Ratio of axis:*

It explains the ratio between the major axis and minor axis of symmetry and calculated through Eq. (3).

$$Ratio\ of\ axis = \frac{Maximum\ axis\ length}{Minimum\ axis\ length} \tag{3}$$

9- *Eccentricity:*

It explains the ratio between maximum chord length and the minimum chord length as in Eq. (4). The maximum chord length or major diameter D of an object O is defined as in Eq. (5)

$$Eccentricity = \frac{Maximum\ chord\ length}{Minimum\ chord\ length} \tag{4}$$

$$D = max_{i,j}\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{5}$$

Where $p_i = (x_i, y_i)$ and $p_j = (x_j, y_j)$ are the pixels in the boundary of the object O.

10- *Extent:*

It explains the ratio of pixels of the region and pixels of the bounding box.

11- *Equivalent Diameter:*

Since a lesion is not a uniform circular object, its diameter was calculated using Eq. (6), as A is the total area of the segmented lesion.

$$Equivalent\ Diameter = \sqrt{\frac{4A}{\pi}} \tag{6}$$

12- *Radius:*

For long lesions, the radius is considered as the half of the equivalent diameter.

$$Radius = \frac{Equivalent\ Diameter}{2} \tag{7}$$

13- *Bounding Box area:*

It represents the area of less rectangle that includes the lesion.

14- *Ratio of bounding box axis:*

It explains the ratio between the X-axis and the Y-axis of the bounding box containing the lesion.

15- *Estimates borderline:*

After tracing the borderline detecting the tumor, it is easy to calculate the size of the traced boundary that can estimate the borderline.

16- *Convex Hull:*

The convex hull is the minimum-bounding polygon around the lesion and can be thought of as a rubber band enclosing the lesion anchored at the edge points.

17- *Convex Hull area:*

It represents the number of pixels in a binary image specifying the convex hull.

18- *Orientation:*

It explains the angle between the x-axis and the major axis of the lesion.

19- *X-axis of bounding box:*

It represents the width of less rectangle that includes the lesion.

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009

20- Y-*axis of bounding box:*

It represents the length of less rectangle that including the lesion.

2.5.2. Texture features

Texture Features are indirectly extracted parameters based on computing local statistical characteristics of pixel intensities. Texture features are calculated by Gray level co-occurrence matrix (GLCM) [29,30] by focusing on the relationship between pixels of different gray levels. GLCM showed pixel with certain intensity i related to another pixel j at a certain distance d and orientation θ.

We selected 21 texture features that included variance and 20 higher order features. The variance which is a first-order statistical feature and the 20 most relevant higher-order texture features are listed below.

1- *Auto correlation:*

$$Auto\ correlation = \sum_{i,j} (i.j)P(i,j) \tag{8}$$

2- *contrast:*

$$Contrast = \sum_{k=0}^{N_{g-1}} K^2 \Big(\sum_{i=1}^{N_g}\sum_{j-1}^{N_g} P_{(i,j)}\Big)^{|i-j|=k} \tag{9}$$

3- *correlation:*

$$Correlation = \frac{1}{\sigma^2}\sum_{i=1}^{N_g}\sum_{j=0}^{g-1}(ij)P_{(i,j)} - \mu^2\left\{i + j - \mu_\chi - \mu_{\mathcal{Y}}\right\}^3 \tag{10}$$

4- *cluster shade:*

$$Cl.shade = \sum_{i=0}^{g-1}\sum_{j=0}^{g-1}\left\{i + j - \mu_\chi - \mu_{\mathcal{Y}}\right\}^3 P(i,j) \tag{11}$$

5- *Cluster Prominence:*

$$Cl.Prom = \sum_{i=0}^{g-1}\sum_{j=0}^{g-1}\left\{i + j - \mu_\chi - \mu_{\mathcal{Y}}\right\}^4 P(i,j) \tag{12}$$

6- *Dissimilarity:*

$$D = \sum_{i,j}^{N-1} P(i,j).|i-j| \tag{13}$$

7- *Energy:*

$$Energy = \sum\nolimits_{i,j} P(i,j)^2 \tag{14}$$

8- *Entropy:*

$$Entropy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{(i,j)} \log\left[P(i,j)\right] \tag{15}$$

9- *Homogeneity:*

$$Homogenity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{P(i,j)}{1 + (i-j)^2} \tag{16}$$

10- *Maximum probability:*

$$Max.propapility = \forall(i,j) max\{P(i,j)\} \tag{17}$$

11- *Sum of squares (Variance):*

$$\sigma^2 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu)^2 P(i,j) \tag{18}$$

12- *Sum average:*

$$Sum\ of\ average = \sum_{k=0}^{2N_{g-2}} K P_{x+y}(K) \tag{19}$$

Where: $P_{x+y} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P_{(i,j)}$

13- *Sum variance:*

$$Sum\ of\ variance = \sum_{k=0}^{2N_{g-2}} (K - f_{sv})^2 P_{x+y}(K) \tag{20}$$

Where: $f_{sv}$is sum of average

14- *Sum entropy:*

$$Sum\ of\ entropy = -\sum_{k=0}^{2N_{g-2}} P_{x+y}(K) \log\left[P_{x+y}(K)\right] \tag{21}$$

Where: $P_{x+y} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P_{(i,j)}$

15- *Difference variance:*

$$Difference\ variance = \sum_{k=0}^{N_{g-1}}\left[P_{|x-y|}(K)\left(K - \sum_{i=0}^{N_{g-1}} iP_{|x-y|}(K)\right)^2\right] \tag{22}$$

16- *Difference entropy:*

$$Difference\ entropy = -\sum_{k=0}^{N_{g-1}} P_{|x-y|}(K) \log\left[P_{|x-y|}(K)\right] \tag{23}$$

17- *Information measure of correlation1:*

$$Information\ of\ correlation1 = \frac{f_{ent} - HXY1}{max\{HX, HY\}} \tag{24}$$

Where: -f_{ent} is entropy

$$-HXY1 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P(i,j) \log\left[P_i, P_j\right] - H = -\sum_{g=1}^{N_g} P_{(i)}P_{(j)} \log\left[P_i P_j\right]$$

18- *Information measure of correlation2:*

$$\text{Information measure of correlation2} = \sqrt{1 - exp[-2|HXY2 - f_{ent}|]} \tag{25}$$

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009

Where: -$f_{ent}$is entropy

$$HXY2 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{(i)}P_{(j)} \log\left[P_iP_j\right]$$

19- *Inverse difference (ID) is human:*

$$ID = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{P(i,j)}{1+|i-j|} \tag{26}$$

20- *Inverse difference normalized (IDN):*

$$IDN = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{P(i,j)}{1+|i-j|/G^2} \tag{27}$$

21- *Inverse difference moment normalized (IDMN):*

$$IDMN = \sum_{i=0}^{N_g}\sum_{j=0}^{N_g} \frac{P(i,j)}{1+(i-j)^2/G^2} \tag{28}$$

Various co-occurrence matrix realizations depending on distance and angle, and displaying different angular relationships, are introduced in our work. The second-order histogram, or grey-level co-occurrence matrix, $H(y_i,y_j,\measuredangle,\theta)$explained the likelihood, distribution of appearance of a pair of grey-level values isolated, by a given displacement d at an angle ofθ. The textural measures were calculated on horizontal, vertical and two diagonal directions o an image as $D_s^{|K|}(i,j)$ is computed for each of the four angels $0^{\hat{A}^\circ}, 45^{\hat{A}^\circ}, 90^{\hat{A}^\circ}$ and$135^{\hat{A}^\circ}$, where K denotes the angel of spatial neighborhood. The geometrical relationships of GLCM measurements made by the four angels under the assumption of angular symmetry. The previously mentioned 20 GLCM features extracted were defined in the four different angels (20 features * 4 different angels) to have for total 80 features.

2.5.3. Moment invariants

In addition, we used moment invariants [31] as features for image processing by calculating seven invariant moments according to image rotation. Moment invariants $\phi_I 1 \le I \le 7$are useful parameters of being unchanged under image scaling, translation and rotation. They are calculated over the shape boundary and are interior region. Applying seven moment invariants on the mammographic image under scaling, translation and rotation is presented and table of computed seven invariant equations is shown in Table 1.

The major objective of feature selection is to minimize features by deleting features with small or no predictive data. In this study, we used Fisher score [32] to decrease the features significantly and increase the detection accuracy. Subsequently, classification is the last step in CAD system where the classifier used the selected robust significant features acquired from the feature selection step as input to classify the abnormal lesion into benign or malignant. The proposed classifiers that used for recognition the classes of the image were KNN [33], ANN and SVM [34]. In this study, two techniques of ANN have proposed, the automatic and traditional technique. Three different SVM approaches have proposed for classification as a result of using three kernel functions linear, quadratic, and radial basis function (RBF) kernels.

3. Results

In this model, CAD technique required for detection of MCC of breast cancer was designed based on scientific fundamentals. Every image was enhanced and segmented to a specified ROI, and different features were extracted, then the most prominent of these features were selected. The resulting selected features were fed into three different classifiers to provide different results based on each feature type, selection method, and classification. Accordingly, different experimental approaches were designed, and their results were discussed and compared to each them. Fig. 4 showed the description of the different techniques which have applied.

3.1. Single models

Single models focused on studying a single type of feature and compare system performance when using one feature selection technique to choose the most significant features that act as input to different classifiers. The GLCM and shape features are proposed as candidate features. Features resulting from feature selection and extraction methods are fed into 6 different classifiers as shown in Fig. 5. We divided signal models into texture analysis approach, geometric feature approach, and invariant moments approach.

3.1.1. Texture analysis approach

Texture features based on GLCM extracted out of the segmented image. 20 GLCM features with the four different angels plus the variance i.e., 81 texture features were extracted from segmented breast cancer images. The implementation of three techniques of preprocessing, the Fisher score as a feature selection and three different methods of classification are shown in Fig. 5.

These features were fed into the different techniques and examined twice, once on (Normal - Abnormal) and (Malignant – Benign). The best accuracy was observed by classifying those features according to ANN, it was 100%. Two different experimental results obtained are mentioned as in Table 2.

Another approach based on texture features is used by GLCM with wavelet transform. 20 GLCM features with the four different angels plus the variance i.e., 81 texture features were extracted from segmented breast cancer images. Accordingly, four techniques of preprocessing including analysis of two-level wavelet transform (3db&6db) to improve texture feature results were implemented.

These features were fed into the different techniques and examined twice, once on (3db wavelet) and once on (6db wavelet). The best accuracy was observed by classifying those features according to SVM and KNN. It was 80% accuracy for SVM and KNN using the 6db wavelet. Two different experimental results obtained are mentioned in Table 3 by different levels of wavelets and different classifiers.

Table 1
Seven Moment Invariants Equations.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7
Half Size	2.1740e + 004	1.7254e + 008	7.4393e + 008	1.4130e + 009	7.0026e + 016	1.8376e + 013	1.2152e + 018
$180^{\hat{A}^\circ}$	8.6822e + 004	2.7475e + 009	2.4037e + 010	4.5332e + 010	5.0848e + 019	2.3347e + 015	1.2318e + 021
$90^{\hat{A}^\circ}$	8.4227e + 004	2.3204e + 009	2.2273e + 010	4.0706e + 010	3.8338e + 020	2.0617e + 015	1.0821e + 021
$45^{\hat{A}^\circ}$	8.1330e + 004	6.6990e + 006	7.8129e + 009	3.5617e + 010	4.5774e + 020	1.9895e + 015	1.44320e + 020

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009

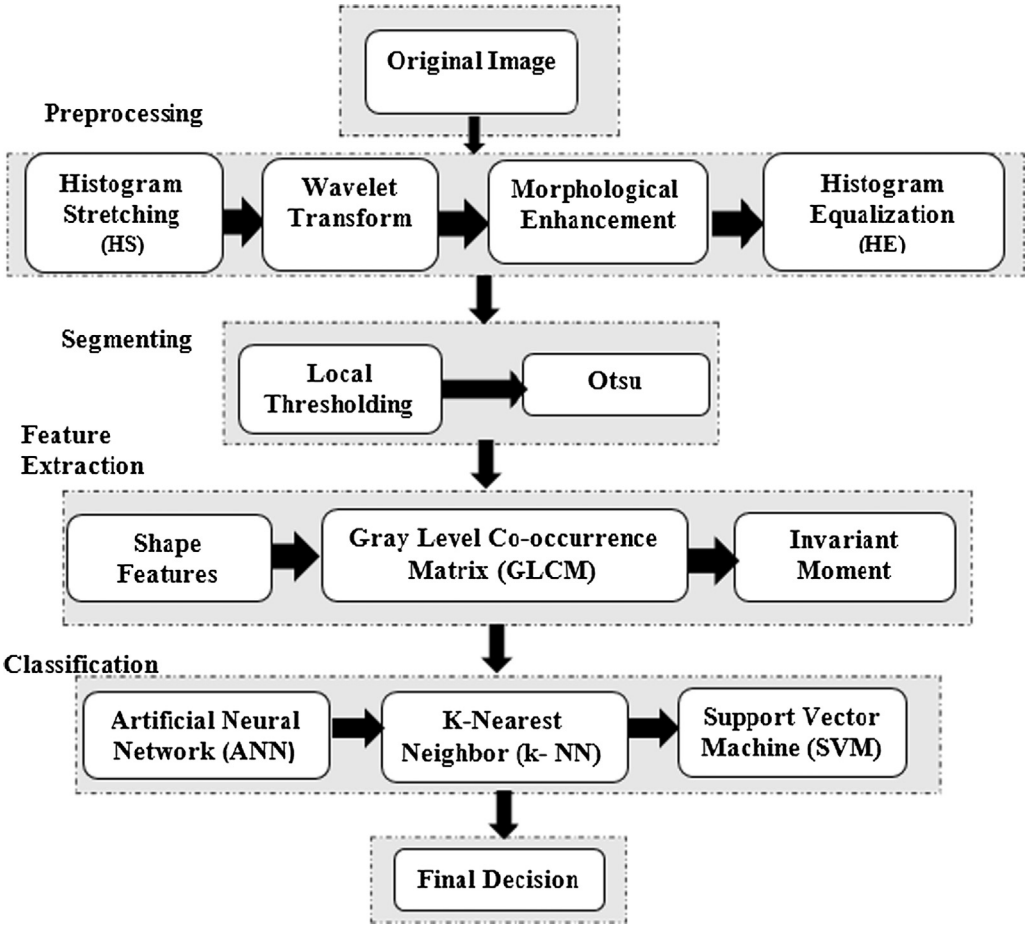

Fig. 4. Single and combined Techniques' approaches.

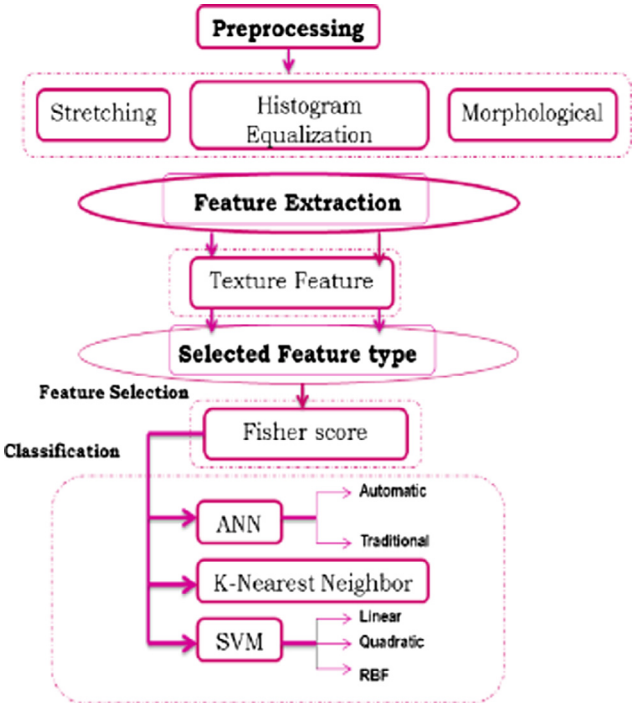

Fig. 5. Block Diagram of Single Model. For texture analysis approach, features are based on GLCM approach. For wavelet transform approach, features are based on GLCM approach with wavelet transform. For geometric approach, features are based on shape features. For invariant moments approach, features are based on invariant moments.

Table 2
Features results for (Normal-Abnormal) and (Malignant-Benign) according to GLCM.

Normal-Abnormal				
	ANN (Automatic)	ANN (Traditional)	KNN	SVM
Sensitivity	100%	100%	78%	70%
Specificity	100%	100%	79%	75%
Accuracy	100%	100%	76%	65%
Malignant-Benign				
Sensitivity	66%	100%	65%	60%
Specificity	77%	77%	75%	85%
Accuracy	71%	61%	70%	70%

Table 3
Features results for (3db Wavelet) and (6db Wavelet) according to GLCM.

3db Wavelet				
	ANN (Automatic)	ANN (Traditional)	KNN	SVM
Sensitivity	100%	66%	77%	63%
Specificity	77%	77%	78%	80%
Accuracy	61%	71%	77%	75%
6db Wavelet				
Sensitivity	100%	66%	93%	95%
Specificity	77%	77%	90%	100%
Accuracy	78%	71%	80%	80%

3.1.2. Geometric feature approach

In this approach, 22 shape features (20 geometric, margin and density) are extracted out of the segmented image. Four techniques of preprocessing are implemented by the Fisher score as feature selection and three different methods of classification.

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009

Table 4
Features results according to Shape features and Invariant Moments.

Shape Features				
	ANN (Automatic)	ANN (Traditional)	KNN	SVM
Sensitivity	60%	65%	70%	65%
Specificity	80%	60%	76%	70%
Accuracy	60%	65%	72%	88%
Invariant Moments				
Sensitivity	100%	100%	70%	83%
Specificity	96%	85%	75%	88%
Accuracy	97%	91%	70%	85%

The best accuracy was observed by classifying those features according to SVM that recorded 88% as mentioned in Table 4.

3.1.3. Invariant moments approach

This approach is based on computing equations for seven invariant moment features which extracted out of the segmented image. We presented the results by using these features according to four techniques of preprocessing, the Fisher score selection and three different methods of classification.

The best accuracy was observed by classifying those features according to ANN. It reached 100% sensitivity and 97% accuracy for automatic mode of ANN. For SVM, the sensitivity was 83% with the accuracy of 85% as mentioned in Table 4.

3.2. Combined models

We divided combined models into dual approaches and multiple approaches. Two combined approaches were suggested as a result of a combination of the best features extracted out of each feature type. Dual approaches were proposed by adding shape features to texture features as in Fig. 6(a).

Twenty-one texture features were added to another twenty-two shape features were introduced to the different classification techniques to evaluate the model performance. It was observed that the combined features which detect all MCC in breast cancer have a high sensitivity of 100%, and the best accuracy of 96% according to the traditional mode of ANN. Table 5 summarized the corresponding results by applying the combined features to the different classifiers.

In multiple approaches, after examining various pairs of feature types together, it was important to examine the accuracy results from using the three different features categories together. This approach based on combined features extractions (texture, shape and invariant moments) that are extracted out of the segmented image. We presented the results of these features that based on four techniques of preprocessing, the Fisher score feature selector and three different methods of classification as shown in Fig. 6(b).

Twenty-one texture, twenty-two shape, and seven invariant moment features were introduced to the six different classification techniques to evaluate the model performance. It was observed that the best results achieved by classifying the three combined features according to the traditional mode of ANN that reached 100% for the sensitivity and 98% for the specificity. Also, best accuracy reached 96% in the automatic mode of ANN. Table 6 showed the results of applying the combined features to the different classifiers.

3.3. Evaluation criteria

Most researchers on the classification of mammogram images were based on the individual type of features extraction. This led to the imposition of restrictions on experimental results. We avoided these restrictions by using more than types of features in order to develop the classification process. We proposed many

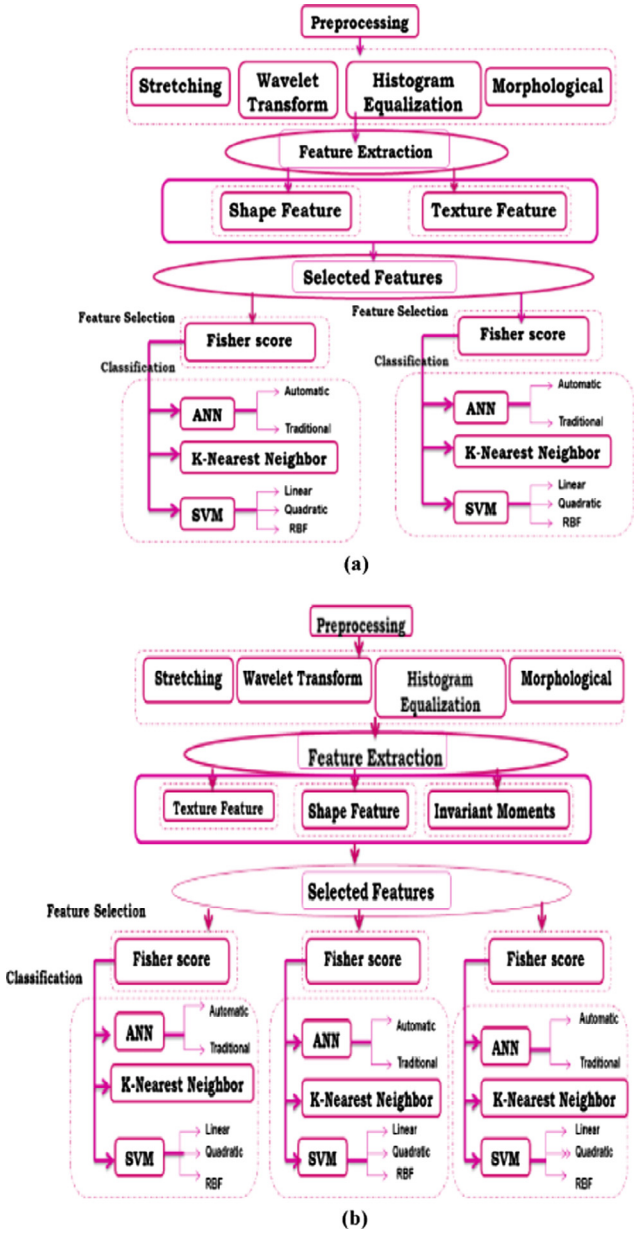

Fig. 6. Block Diagram of (a) Dual Model and (b) Combined Model.

Table 5
Classification results for shape and Texture features.

	ANN (Automatic)	ANN (Traditional)	KNN	SVM
Sensitivity	100%	66%	93%	95%
Specificity	77%	77%	90%	100%
Accuracy	78%	71%	80%	80%

trials for choosing the best features sets. Great controversy of features selection will remain a tool for motivation the work in pattern recognition.

4. Discussions

Radiologists used CAD system as a second reader for the mammogram to improve the performance of cancer detection [35] due

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009

Table 6
Classification results for shape, Texture and Invariant Moment features.

	ANN (Automatic)	ANN (Traditional)	KNN	SVM
Sensitivity	98%	100%	65%	65%
Specificity	94%	98%	80%	90%
Accuracy	96%	92%	77%	75%

to the dependence on radiologist decisions only is not enough for helping patients. Because of the lack of human resource, radiologists compared the CAD's results with their first opinion, and then they mark these abnormalities in ROI, which is used as input to CAD. This paper is based on establishing a model for a CAD system that can be so beneficial for the radiologist in recognizing breast cancer changes with premature stage and quicker than traditional examination programs.

This proposed CAD system was built for the detection and classification of masses. The system was focused on the detection of MCC and the extraction of the ROI for accurate segmentation, then the extraction of features sets from mammogram images, which were used in the ranking of breast cancer mutations by means of three classifiers.

The system depended on mammographic enhancement using four techniques, namely histogram stretching, wavelet transform, morphological analysis, and HE. All obtained results were very supportive and indicated that the suggested system is a successful technique to elicit MCC and develop them, as they are established in irregular breast tissues. This is designed automatically to get more accurate results and encourage the radiologists' treatment of breast cancer at an early phase.

The majority of previous experiments on breast cancer detection addressed the problem of features selection and competed for improving the diagnosis efficiency. Therefore, results of mammogram classification based on image preprocessing algorithms and selection of the best feature sets. Recently, researchers anticipated deep learning approaches [36,37] on mammogram images for discrimination of breast cancer and compared results with traditional classification techniques. Although, the power of deep learning approaches that based on implementation without features selection, features selection is still freely able to control results. Also, transfer learning approaches [38] are applied on mammogram image to raise the power of CAD system with an accuracy of 97.50%.

The suggested model proposed in this work achieved the better results rather than previous similar works in [39–41]. Algorithms in [39] used morphological enhancement, HE, followed by segmentation based on Otsu's threshold for the identification of MCC, and at last classification stage that recorded the accuracy of 73% by KNN and 77% by ANN. Recent work [40] used Radial Basis Function Neural Network (RBFNN) for mammograms classification based on GLCM using texture-based features, where RBFNN's accuracy was 93.98%.

We considered that first phase of our work discussed in [39] and the second phase presented in this paper to improve results than in the first phase. The combined model is shown to have a major possibility for breast cancer diagnosis in digital mammograms, more flexibility and hence better diagnostic accuracy rather than a dual model. Based on the good results of this study, we suggested rebuilding an online diagnosing system in the future and using 3D mammograms structures as the next phase.

5. Conclusions

Generally, breast cancer detection and diagnosis have developed to underpin image processing technology. Thus, image processing techniques have been prospered in mammograms field to avoid the errors of radiologist's scanning of breast images. CAD system is used to ameliorate the reading of mammographic images by contrast enhancement, segmentation, merged features and classifications techniques for MCCs detection in breast cancer. Ultimately, the proposed model for early detection and diagnosis of breast cancer that based on a combination of different features clarified the declaration of revolution in the advancement of healthcare. This model is fully automated to get more accurate results and facilitate the radiologists' diagnosis of breast cancer at an early stage.

References

[1] Zheng Bin, Hardesty Lara A, Poller William R, Sumkin Jules H, Golla Sara. Mammography with computer-aided detection: reproducibility assessment - initial experience. Radiology 2003;228(1):58–62.
[2] American Cancer Society [https://www.cancer.org/]. Cancer Facts and Figures 2006. Atlanta.
[3] Gurcan Metin N, Chang Heang_Ping, Sahiner Berkman, Hadjiiskil Lubomir, Petrick Nicholas, Helvie Mark A. Optimal neural network architecture selection: Improvement in computerized detection of microcalcifications. AcadRaiol 2002;9:420–9.
[4] Tartar Marie, Comstock Christopher E, Kipper Michael S. Breast Cancer Imaging: A Multidisciplinary, Multimodal Approach. Mosby Elsevier; 2008.
[5] Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. Radiology 2001;220:781–6.
[6] Vyborny CJ. Can computers help radiologists read mammograms? Radiology 1994;191:315–7.
[7] Doi Kunio. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph. 2007;31(4–5):198–211.
[8] Stefanoyiannis AP, Costaridou L, Sakellaropoulos P, Panayiotakis G. A digital density equalization technique to improve visualization of breast periphery in mammography. Br J Radiol 2000;73:410–20.
[9] Byng JW, Critten JP, Yaffe MJ. Thickness-equalization processing for mammographic images. Radiol 1997;203:564–8.
[10] Pohlman S, Powell K, Obuchowski N, Chilcote W, Grundfest Broniatowski S. Quantitative classification of breast tumor in digitized mammograms. Med Phys 1996;23(8):1337–45.
[11] Petrick N, Chan H, Sahiner B, Wei D. An adaptive density weighted contrast enhancement filter for mammographic breast mass detection. IEEE Trans Med Imaging 1996;15:59–67.
[12] Kegelmeyer Jr WP, Pruneda JM, Bourland PD, Hillis A, Riggs MW, Nipper ML. Computer-aided mammographic screening for speculated lesions. Radiology 1994;191:331–7.
[13] Laws K. Textured image segmentation. University of Southern California; 1980. Ph.D. Thesis.
[14] Suckling J et al. The Mammographic Image Analysis Society Digital Mammogram Database. Excerpta Medica Int Congress Series 1994;1069:375–8.
[15] Heath Michael, Bowyer Kevin, Kopans Daniel, Moore Richard, Kegelmeyer WPhilip. The Digital Database for Screening Mammography. Medical Physics Publishing; 2001. p. 212–8.
[16] Bovik A. Handbook of Image and Video Processing. San Diego: Academic Press; 2000.
[17] Mousa R, Munib Q, Moussa A. Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. Expert Syst Appl 2005;28:713–23.
[18] Nakayama R, Uchiyama Y, Yamamoto K, Watanabe R, Namba K. Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms. IEEE Trans Biomed Eng 2006;53(2):273–83.
[19] Sivarajan U, Jayapragasam KJ, Abdul Aziz YF, Rahmat K. Dynamic contrast enhancement magnetic resonance imaging evaluation of breast lesions: a morphological and quantitive analysis. JHkcoll Radiol 2009.
[20] Rizzi M, D'Aloia M, Castagnolo B. Computer aided detection of microcalcifications in digital mammograms adopting a wavelet decomposition. Integr. Comput.-Aid. E. 2009;16:91–103.
[21] Strickland RN, Hahn HL. Wavelet transforms for detecting microcalcifications in mammograms. IEEE Trans. Med. Imag. 1996;15(2):218–29.
[22] Wirth Michael A, Stapinski Alexei. Segmentation of the breast region in mammograms using snakes. IEEE Proceedings of First Canadian Conference on Computer and Robot Vision, 2004. London, Ontario, Canada.
[23] Brake GM, Karssemeijer N. Segmentation of suspicious densities. Med. Phys. 2001;28:258–66.
[24] Kai Xu, Kun Qin, Tao Pei. Interactive method for image segmentation based on cloud model. Comput Eng Appl 2006;34:33–5.
[25] Jeong HJ, Kim TY, Hwang HG, et al. Comparison of thresholding methods for breast tumor cell segmentation. In: IEEE Proceedings of 7th International Workshop on Enterprise Networking and Computing in Healthcare Industry. p. 392–5.
[26] Saha PK, Udupa JK. Optimum threshold selection using class uncertainty and region homogeneity. IEEE Trans Pattern Anal Mach Intell 2001;23:689–706.

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009

[27] Thangavel K, Karnan M. Computer aided diagnosis in digital mammograms: detection of microcalcifications by meta heuristic algorithms. GVIP J 2005;5 (7).
[28] Kwok Rchandrashekar SM, Attikiouzel Y. Automatic Pectoral Muscle Segmentation on Mammograms by Straight Line Estimation and Cliff Detection. In: 7th Australian an New Zealand Intelligent Information Systems Conference. Perth, Western Australia. p. 18–21.
[29] Mirzaalian H, Ahmedzadeh MR, Sadri S. Pectoral muscle segmentation on digital mammograms by nonlinear diffusion filtering. IEEE Pacific Rim Conference Commun Comput Signal Processing 2007:581–4.
[30] Haralick Robert M, Shanmugam K Dinstein. Textural Features for Image Classification. IEEE Trans Syst Man Cybernet 1973;3(6):610–21.
[31] DanielMadan Raja S, Shanmugam A. Artificial neural networks based war scene classification using invariant moments and GLCM features: a comparative study. Int J Eng Sci Technol 2001;3(2).
[32] Duda PEHRO, Stork DG. Pattern Classification. Wiley-Interscience Publication; 2001.
[33] Panigrahi BK, Pandi VR. Optimal feature selection for classification of power quality disturbances using wavelet packet-based fuzzy k-nearest neighbor algorithm. Generation Transmission Distribut 2009;3:296–306.
[34] Thai Le Hoang, Hai Tran Son, Thuy Nguyen Thanh. Image classification using support vector machine and artificial neural network. In: I.J. Information Technology and Computer Science. p. 32–8.
[35] Kaymaka Sertan, Helwan Abdulkader, Uzun Dilber. Breast cancer image classification using artificial neural networks. Procedia Comput Sci 2017;120:126–31.
[36] Kallenberg M, Petersen K, Nielsen M, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans Med Imaging 2016;35(5):1322–31.
[37] Abdel-Zaher AM, Eldeib AM. Breast cancer classification using deep belief networks. Expert Syst Appl 2016;46:139–44.
[38] Vesal Sulaiman, Ravikumar Nishant, Davari AmirAbbas, Ellmann Stephan, Maier Andreas. Classification of breast cancer histology images using transfer learning. In: 15th International Conference on Image Analysis and Recognition. p. 812–9.
[39] Mohamed Hayat, Mabrouk Mai S, Sharawy Amr. Computer Aided Detection System for Microcalcifications in Digital Mammogram. Comput Methods Programs Biomed 2014;116(3):226–35.
[40] Pratiwi Mellisa, Alexander Jeklin Harefa, Nanda Sakka. Mammograms classification using gray-level co-occurrence matrix and radial basis function neural network. In: International Conference on Computer Science and Computational Intelligence (ICCSCI). p. 83–91.
[41] Komura Daisuke, Ishikawa Shumpei. Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J 2018;16:34–42.

Mai S. Mabrouk received her B.Sc., M.Sc. and phD degrees from the Biomedical Engineering Department at Cairo University in 2000, 2004 and 2008 respectively. She is currently an Associate professor of and department head of Biomedical Engineering at Misr University for Science and Technology. Her biography was selected to appear in Marquis Who's Who in the World in 2012. Along her career, she was a technical reviewer and editorial board member for several international journals and conferences. She published over 70 peer-reviewed journal and conference articles in the areas of medical imaging processing, Bioinformatics and human computer interface.

Please cite this article as: M. S. Mabrouk, H. M. Afify and S. Y. Marzouk, Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2019.01.009