




Research Article

Multi-mass breast cancer classification based on hybrid descriptors and memetic meta-heuristic learning



Khosro Rezaee¹  · Adele Rezaee² · Negar Shaikhi³ · Javad Haddadnia⁴

Received: 15 November 2019 / Accepted: 20 May 2020 / Published online: 27 June 2020
© Springer Nature Switzerland AG 2020

Abstract

Among the causes of death in the world, breast cancer is considered the most common cause of mortality among women to the extent that one in five deaths among women is attributed to the incidence of this cancer. In this paper, we introduce a computer-aided detection approach to multiple classifications of breast masses. We tried to separate and intelligently recognize different masses in the breast cancer by means of mammograms so that in the first step, with the pre-processing, pectoral region is segmented from other parts and different areas are primarily clustered by *K*-means method. In the next step, using aggregation of efficient features such as texture features, Pseudo-Zernike moments, and wavelet features will be extracted from the input image and simulated annealing algorithm will reduce the size of feature vector. The final step will be the classification of possible masses in mammogram and the assessment of its severity based on memetic meta-heuristic adaptive neuro-based fuzzy inference system in which the optimizer is shuffled frog-leaping algorithm. The proposed method is evaluated using 322 mammogram images taken from Mini-MIAS database, which contain a variety of possible masses in mammograms. We compare our model with similar algorithms and several state-of-the-art methods through a comprehensive set of experiments. In this approach, the focus is on providing a hybrid algorithm for accurate detection and extraction of masses in mammography, with the approach that the physician can predict both the potential disease stage and type of tumor.

Keywords Breast cancer · Mammograms · Feature extraction · Simulated annealing · Optimal ANFIS

1 Introduction

Although cancer in some cases includes benign tumors, there is also the possibility of malignant tumors and hence great increase in the rate of mortality [1, 2]. One of the cancers in women which causes high rate of mortality as a result of the malignant masses is breast cancer [3]. In some European, Africa and Asia countries, the rate of mortality caused by this disease is increasing [4] and according to 2011 statistics, 110 women die every day from breast cancer globally [5]. Studies have shown that the prevention of this disease as a result of unknown factors seems very

complicated, but in the early stages of formation, diagnostic process can be applied [6]. So, early detection and diagnosis is one of the most important factors in the treatment of this disease. Breast cancer is the leading cause of mortality among women population and is responsible for one-fifth of all deaths [7]. For instance, the number of patients with breast cancer has been on the increase in Asian countries and the age of disease onset is 10 years less than Western countries [8]. What is obvious is that this cancer, especially among women, is very common and dangerous and urged researchers and physicians to look for ways to identify and harness this wave of cancer. On the

✉ Khosro Rezaee, Kh.rezaee@meybod.ac.ir | ¹Department of Biomedical Engineering, Meybod University, Meybod, Iran. ²Department of Biomedical Engineering, Islamic Azad University, Tabriz Branch, Tabriz, Iran. ³Department of Engineering, Faculty of Nuclear Engineering, Central Tehran Branch, Azad University, Tehran, Iran. ⁴Department of Biomedical Engineering, Hakim Sabzevari University, Sabzevar, Iran.



other hand, for an early diagnosis of this disease, the existence of an intelligent system with high accuracy for detection of cancerous masses is highly important. This cancer is usually diagnosed via surgical biopsy which has the higher accuracy among the existing method, but the difference is that this method is an invasive, time-consuming and expensive procedure [9]. Mammography is currently the most common and popular method for early diagnosis of this disease which has decreased the mortality rate to 25% due to early detection [10]; nevertheless, the interpretation of images resulting from mammography is very difficult and according to official figures of the National Cancer Institute in the US, 10–30% of glands in patient's breast in mammography images are indistinguishable by radiologists [11]. Also in the mammography method, 30% of breast cancers due to the lack of precise detection of mass locations are not recognized properly [12, 13]. Therefore, employing computer-aided diagnosis (CAD), in the field of mammography can be useful for more accurate interpretations by specialist. CAD can be specifically helpful in intelligent detection of diseases from medical images. As seen in previously proposed systems and other studies, identification employing powerful extraction CADs and selection of features as well as classification in diagnosis of cancer masses included better results [14–16]. Using methods based on image processing can greatly increase the chance of detection in mammography [17]. Overall, the utilization of the proposed CAD system will lead to 80–90% detection accuracy [18].

The problem appears to be that most previous methods in this field only identify the presence or absence of tumor and as a result, only few researches has been carried out on an automated approach for recognizing the masses in mammography images [19]. The segmentation procedure in mammography images is one of important problem that is vital step for gathering information from the masses [20, 21].

Some researchers have also worked exclusively on two types of benign and malignant or micro-calcification masses. But in most cases, these methods have taken the advantages of intelligent systems model in inference and recognition of appropriate patterns with the possibility of learning [22]. A substantial number of former methods have benefited from tissue analysis or shape-based attributes [23–33].

Efficient tools that have been employed in various fields of image analysis have also extracted various features in the field of mammographic image processing including local binary pattern [34], gabor features [35, 36], histogram [37], principle component analysis [38] and geometric and statistical characteristics [31]. In a number of methods, Zernike moment method has been employed for the description and extraction of features [22, 39–41].

Using more features extracted from other tools like texture and shape characteristics can greatly increase the accurate identification of tumor type. Various studies [23–25, 28] have utilized texture features alongside non-texture features. Kabbadj et al. [31] employed geometric and statistical features, while Beura et al. [42] employed the wavelet transform along with gray-level co-occurrence matrix (GLCM) in the diagnosis of masses.

A number of other studies also tried to optimize sample classification like Singh et al. [22] that made employed adaptive differential evolution wavelet neural network (Ada-DEWNN) model, which was an optimized model. Also, Dheeba et al. [5, 35] and Raghavendra et al. [43] applied neural network optimized model and chose Gabor filters as the features extraction tool for mammography image. Recently, the conventional neural network is used as the core of an integrated belief concept for dealing with the assortment problem or feature extraction of the breast lesion detection and classification [44–47]. Also, Xie et al. [48] used extreme learning machine (ELM) as powerful classifier for breast mass classification in digital mammography. When the comparison is based on the use of classification type, *k*-NN [26, 27, 49], support vector machine [25, 29–31, 34, 36, 38, 40], artificial neural network [22, 32, 35, 39, 41–43], fuzzy inference system [23] and other efficient classifiers like ANFIS [50] are among those that have been frequently used. Despite the desired accuracy in their study and the small dimensions of extracted texture features, the selection method, the number of masses and their type in recognition is ambiguous. In addition to the data obtained by researchers, there are some databases such as MIAS, DDSM, DBT and IRMA in this field which are used by researchers for data analysis. The number of disease classes in images has been mentioned in some of these databases.

Other researchers have exclusively explored other masses like micro-calcification [29, 31, 51, 52] and some other researchers alongside images without symptoms; have been trying to recognize benign and malignant tumors [42]. The number of classes in a mammography image may exceed 10 types of masses; for instance, in the DDSM database, there are about 12 different classes of benign and malignant and similar masses, while in the Mini-MIAS database, the maximum number of masses does not exceed 7 [42]. In addition, there are some tools for assessment of the research such as the calculation of accuracy, sensitivity and specificity that can be considered appropriate benchmarks.

In this paper, we address a framework to detect the multi-mass breast cancer based on hybrid descriptors and memetic meta-heuristic learning. The novelty of our study is the analysis of mammography images using hybrid descriptors such as Pseudo-Zernike moment and

wavelet transform. Furthermore, we optimize ANFIS classifier based on Memtic shuffled frog-leaping algorithm (SFLA). The remaining part of this paper is organized as follows. The framework of proposed algorithm will be presented in Sect. 2. In Sect. 4, the experimental results of the simulation will be presented and in the same section, the results are compared with other similar methods. Finally, overall conclusion of the system performance will be presented in Sect. 5.

2 Overview of the proposed system

Implementation steps include applying some basic steps in pre-processing of mammography input image, extraction and selection of the best features from the set of aggregated features and finally, the classification based on the ANFIS model. The suggested steps are shown in Fig. 1.

2.1 Pre-processing step

Pre-processing steps comprise three basic steps as follows: (a) removing redundant information from mammography image, (b) deletion of pectoral from breast, and (c) separation of masses using K-means clustering.

2.1.1 Pectoral muscle

In the pre-processing section, simulation and mapping masks in pervious methods [53–55] were utilized for its higher accuracy in the separation of images especially from MIAS database at the beginning, probability density function used for the allocation of any part belonging to the image and area, were divided into three sections of background, breast and pectoral muscle.

$$A_R(x) = \frac{n(x \in R)}{N} \tag{1}$$

where A represents the probability density function for each pixel position x and degree of belonging to the area of R . On the other hand, $n(x \in R)$ is the number of x positions in the area of R . Furthermore, N is assumed as the total number of analyzed images. In order to define the probability density function, light intensity information is used to create processed masks [55].

$$I_R(i) = \frac{H_R(i)}{\sum_{j=1}^3 H_j(i)} \tag{2}$$

where I_R refers to the probability of any light intensity in the area of R and H is the intensity histogram. Finally, the label is assigned to each pixel by correspondence between LBP codes and histogram and also computed as the texture descriptor of that region. Therefore, it is possible to assign probabilities to each LBP code associated with the three regions of the tissue [55].

$$T_R(t) = \frac{LBR_R(i)}{\sum_{j=1}^3 LBP_j(i)} \tag{3}$$

T refers to tissue information and here, LBP histogram is related to the code listed and thus the final data in order to build a probabilistic model for separation is mentioned [55]:

$$P_R(p(x, i, t)) = A_R(x)I_R(i)T_R(t) \tag{4}$$

Eventually, pectoral segmentation was realized based on the logical operator method with initial masks that had been defined manually by radiologists. According to logical operators of image processing, the AND operator is applied with the corresponding original mammogram and the pectoral area becomes segmented.

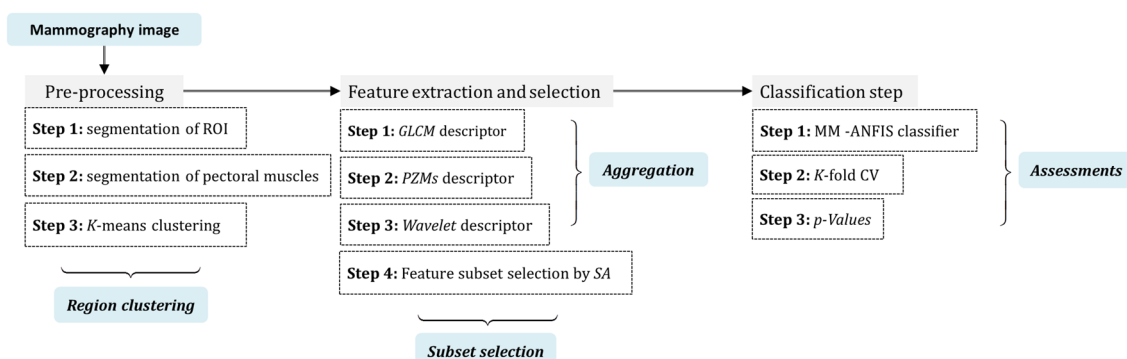


Fig. 1 Block diagram of the proposed scheme for multi-classification of mammograms

2.1.2 Region clustering

After eliminating redundant information, the obtained image is clustered using K-means method. By selecting the appropriate cluster or clusters, the masses in mammography image can be separated. If input patterns include a set of N vectors $\{ \}$ and the Euclidian distance is used as a measure of similarity, then we can formulate K-means clustering as that of finding K cluster centers, that minimize the total square-error E [56]:

$$E(\vec{c}_1, \vec{c}_2, \dots, \vec{c}_K) = \sum_{k=1}^K \left\{ (1/N) \sum_{i=1}^N m_{ki} \|\vec{x}_i - \vec{c}_k\|^2 \right\} \tag{5}$$

where $m_{ki} = 1$ if \vec{x}_i belongs to cluster k , and $m_{ki} = 0$ otherwise. The notation $\| \cdot \|$ denotes norm of term. When the training patterns are generated from probability density $p(\vec{x}_i)$ defined on an input space S , the cost function of the K-means algorithm is transformed into:

$$E(\vec{c}_1, \vec{c}_2, \dots, \vec{c}_K) = \sum_{k=1}^K \int_S m(\vec{x}) \|\vec{x}_i - \vec{c}_k\|^2 p(\vec{x}) d\vec{x} \tag{6}$$

where $m(\vec{x}) = 1$ if \vec{x}_i belongs to cluster k , and $m(\vec{x}) = 0$ otherwise. For expectation maximization and standard k -means algorithms, the Forgy method of initialization is preferable. Based on this clustering, pixels can be divided into a maximum of 255 clusters. Here, based on the results achieved from Salvador, the number of proposed clusters has been 4–7 [57].

2.2 Feature aggregation

The features are created from the aggregation of extracted features by several tissues and statistical descriptors that have a desired effect on the accuracy of diagnosis. These features are composed of three parts.

2.2.1 Texture features

The GLCM is a square matrix whose elements correspond to the relative frequency of occurrence of a pair of gray values at a certain distance and a determined direction. The elements of a co-occurrence matrix with dimensions of $G \times G$ and distance vector d ($dx = dy$) are defined as (7):

$$P_d(i, j) = \{ ((r, s), (t, v)) : I(r, s) = i, I(t, v) = j \} \tag{7}$$

where $I(\dots)$ represents image with dimensions $N \times N$ and the gray level G . GLCM is in fact the description of P_{ij} frequencies that have two neighboring pixel with distance d , one with the gray intensity i and the other with gray intensity j , that occur within a given neighborhood in the Image. Therefore, GLCM will be formed by a square matrix

whose size depends on the maximum intensity of the gray pixels in the image. Each P_{ij} element represents the number of events of the above structure: pixel with size i in a determined distance d from the pixel j . If $d = 1$, four possible orientations are the possible angles between two pixels can be defined by 0, 45, 90 and 135°, according to Fig. 2.

2.2.2 Pseudo-Zernike moments (PZMs)

PZMs are employed to extract features that do not change with dataflow, that are non-repetitive and are resistant to the noise and visual form of investigated image. However, the most striking feature is the multistage display ability of this technique [58]. The Zernike polynomials are a set of orthogonal polynomials that arise within a unit circle ($x^2 + y^2 = 1$) and is displayed with $V_{nm}(x, y)$ and its structure is defined in (8):

$$V_{nm}(x, y) = V(\rho, \theta) = R_{nm}(\rho) e^{im\theta} \tag{8}$$

In this equation, $j = \sqrt{-1}$, $\theta = \tan^{-1} \left(\frac{y}{x} \right)$, $|\rho| \leq 1$, $n \geq 0$, $m \leq n$ and $n - |m| = \text{even}$. It is also worthy of note that ρ is assumed to be the length of the vector origin to point (x, y) while θ is the angle between vector ρ , and the x-axis in an anti-clockwise direction. As previously mentioned in the above relation, n is the non-negative integer that shows the order of polynomial. The order of horizontal arc and its absolute value is less than or equal to n ($\leq n$), and the difference of m from n is always even. On the other hand, R_{nm} is a radial polynomial that is calculated according to (9) and (10) [40, 41]:

$$R_{nm}(x, y) = \left[\sum_{s=0}^{n-|m|} S_{n,|m|,s}(x^2 + y^2) \right]^{\frac{n-s}{2}} \tag{9}$$

$$S_{n,|m|,s} = (-1)^s \frac{(2n + 1 - s)!}{s!(n + |m| - s)!(n - |m| - s + 1)!} \tag{10}$$

Zernike moments (ZMs) are images mapped into a set of Zernike mixed polynomials. One of the important features of Zernike moment is their orthogonal property, therefore

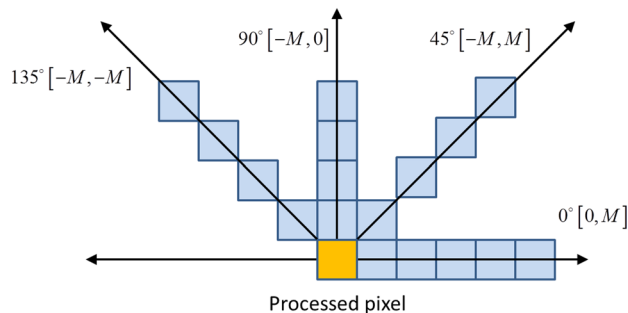


Fig. 2 Feature extraction by using different angles in GLCM

image features without any redundancy of information or overlap between the moments can be presented. Mixed Pseudo-Zernike moments with order n and repetition m are calculated using (11):

$$PZM_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x,y) V_{nm}^*(x,y) \tag{11}$$

where $f(x, y)$ represents the brightness intensity function of the digital mammography image at x and y locations and symbols $*$ also refer to the complex conjugate. Furthermore, it should be noted that the pixels of any image that fall outside the unit circle after mapping will not be utilized in calculating Zernike. The Pseudo Zernike polynomials in a unit circle are shown in Fig. 3.

Furthermore, the Pseudo-Zernike moments for a digital image with dimension $N \times N$ when $0 \geq \rho_n \geq 1$ is displayed according to (12):

$$\begin{aligned} PZM_{nm} &= \frac{n+1}{\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(r,c) V_{nm}^*(r,c) \\ &= \frac{n+1}{\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(r,c) R_{nm}(\rho_{rc}) e^{-jm\theta_{rc}} \end{aligned} \tag{12}$$

2.2.3 Wavelet transform

The process of decomposing multiple signals ($x[n]$) after mapping is carried out with two filters. Each step of the process includes two digital and sampling filters by a factor of 2. In the first filter, $g [.]$ is the discrete wavelet and inherently high-pass while $h [.]$ is the mirror versions of

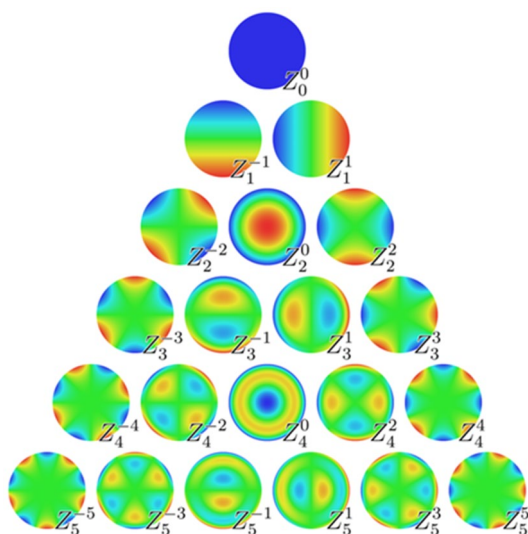


Fig. 3 The first 21 Zernike polynomials, ordered vertically by radial degree and horizontally by azimuthal degree

the wavelet which are inherently low-pass. The first time sampled output signal for high-pass and low pass filters includes partial coefficients D_1 factors and approximation coefficients A_1 , respectively. A_1 is the first approximation coefficients that decompose more than any other factors. All wavelet transforms can be determined in the form of a low-pass filter in the (13):

$$H(z)H(z^{-1}) + H(-z)H(-z^{-1}) = 1 \tag{13}$$

where $H(z)$ is the function z of filter h , and complementary high-pass filter could then be stated on the (15):

$$G(z) = zH(-z^{-1}) \tag{14}$$

A series of filters with increasing length (with index i) can be obtained according to (16):

$$\begin{aligned} H_{i+1}(z) &= H(z^{2^i})H_i(z) \\ G_{i+1}(z) &= G(z^{2^i})H_i(z) \end{aligned} \quad i = 0, \dots, l-1 \tag{15}$$

where $H_0(z) = 1$ is assumed to be the original condition, and two-scale relationship in the time domain can be expressed on the basis of relations (17):

$$\begin{aligned} h_{i+1}(k) &= [h]_{\uparrow 2^i} * h_i(k) \\ g_{i+1}(k) &= [g]_{\uparrow 2^i} * h_i(k) \end{aligned} \tag{16}$$

where $[.]_{\uparrow 2^i}$ represents upward sampling with a factor m and k is assumed as the discrete sampling time. Basic functions and normalized wavelet $\psi_{i,l}(k)$ and $\varphi_{i,l}(k)$ can be defined in the following form:

$$\begin{aligned} \varphi_{i,l}(k) &= 2^{i/2} h_i(k - 2^i l) \\ \psi_{i,l}(k) &= 2^{i/2} g_i(k - 2^i l) \end{aligned} \tag{17}$$

where $2^{i/2}$ results from normalized inner product; i and l are parameters of scale and translation, respectively. The decomposition of discrete wavelet transform is expressed in (19):

$$\begin{aligned} a_{(i)}(l) &= x(k) * \varphi_{i,l}(k) \\ d_{(i)}(l) &= x(k) * \psi_{i,l}(k) \end{aligned} \tag{18}$$

where $a_{(i)}(l)$ and $d_{(i)}(l)$ represent the approximate coefficients and partial coefficients in attention i [59]. Due to this calculation, we are able to practically decompose and subsequently reconstruct the signal. After applying the conversion on the audio signal input from heart, statistical characteristics will be available for the distribution of time-frequency domain.

2.3 Feature subset selection

The basic premise of using feature subset selection algorithms is that the set of extracted data contains both

redundant information and irrelevant features and thus, this process is implemented without incurring much loss of information. Heuristic algorithms belong to the set of powerful techniques to both redundant information elimination and irrelevant features that could be used in optimized feature subset selection in accordance with the resulted error of applying cost function based on unsupervised classifiers. Simulated annealing (SA) is one of large space searching algorithm that is defined as a probabilistic technique for approximating the global optimum of a given function.

In the process of refrigeration, the metals are heated to a high temperature and thereafter, a gradual cooling and reducing of temperature is carried out on them. In this process, an increase in temperature of the metal leads to an increase in speed in the movement of atoms and then a gradual decrease in temperature caused the formation of certain patterns in the position of the atoms. We applied this property to find optimum solutions or the best aggregated features. The best features are found based on the cost in

cost function and finding minimum error of neural network classification. Generally, the process will be as follows:

1. Choosing a random feature subset for search and fitting by neural network
2. Setting the temperature to start
3. 3 Producing a new point to achieve efficient feature subset
4. Evaluating the new Point to accept or reject it as an optimal feature
5. If produced feature subset was better than the first feature subset, they are accepted; otherwise they are accepted with a probability that depends on the temperature and energy in two modes.
6. Temperature drops and steps 3–6 continued to reach the minimum temperature.

The steps for choosing the best properties among the aggregated characters in the algorithm are shown in Algorithm 1.

Algorithm 1. The algorithm to efficient feature subset selection by Simulated Annealing (SA)

Input: The aggregated features

Output: Features index

1: **Begin**

2: **Initialize** $S_{best} \leftarrow S$, $p_{best} \leftarrow p$, $k \leftarrow 0$, $k_{max} \leftarrow$ Constant value (e.g. ∞), $MaxScore \leftarrow$ A constant value

3: **Evaluation count**

4: **While** $k < k_{max}$ and $p < = MaxScore$

5: **Do** $S_{new} \leftarrow$ Neighbor(s)

6: $p_{new} \leftarrow P(S_{new})$

7: **If** $\exp(p_{new} - p) > \text{Rand}()$

8: **Then**

9: $S \leftarrow S_{new}$, $p \leftarrow p_{new}$

10: **If** $p_{new} > p_{best}$

11: **Then**

12: $S_{best} \leftarrow S_{new}$, $p_{best} \leftarrow p_{new}$

13: **End**

14: **End**

15: $k \leftarrow k + 1$

16: **End**

17: **Return** S_{best} , p_{best}

18: **End**

2.4 Classification step

One of the efficient tools in identifying the association between variables is the ANFIS approach that has a similar structure to neural networks and fuzzy systems and it is similar to the neural networks in terms of structure and configuration. ANFIS training is carried out using two algorithms of back-propagation algorithm or combinatorial algorithm including two least squares estimation of the error and back-propagation error which estimate fuzzy membership function parameters. Assuming that the fuzzy system has two inputs x and y and output is z , then the rules are written as shown in (19):

$$\begin{aligned}
 &\underbrace{\text{If } x \text{ is } A_1 \text{ and } y \text{ is } B_1}_{\text{Rule 1}} \longrightarrow \text{Then } f_1 = P_1x + Q_1y \\
 &\underbrace{\text{If } x \text{ is } A_2 \text{ and } y \text{ is } B_2}_{\text{Rule 2}} \longrightarrow \text{Then } f_2 = P_2x + Q_2y
 \end{aligned}
 \tag{19}$$

And if the mean center of defuzzification is to be used for defuzzification, then the output is as follows:

$$\begin{aligned}
 f &= \frac{w_1f_1 + w_2f_2}{w_1 + w_2} = \bar{w}_1f_1 + \bar{w}_2f_2 \\
 \bar{w}_1 &= \frac{w_1}{w_1 + w_2}, \quad \bar{w}_2 = \frac{w_2}{w_1 + w_2}
 \end{aligned}
 \tag{20}$$

2.5 Shuffled frog-leaping algorithm (SFLA)

In the SFLA optimization algorithm, rousing the idea of the frog movement, a strategy is proposed to scan for the parameters improvement, whose adequacy in finding a solution is considerable, compared with different responses. In fact, using this optimization procedure, an ANFIS structure is found to have the least amount of mean square error (MSE) in finding the network output compounds. In other hand, a configuration with neural network weights can be found that could lead to a best classification with a negligible error.

Deciding these parameters will incredibly influence the specified exactness. To discover the leading structure of ANFIS, we propose that the SFLA algorithm perform the optimization. The steps of SFLA to discover best parameters of ANFIS classifier are as follows [60]:

Step 1: Initialization H frogs are randomly generated to construct the initial population. The position of the h_{th} frog is encoded as $X_h = [x_{h1}, x_{h2}, \dots, x_{hd}, \dots, x_{hD}]$, $h = 1, \dots, H$, whose, D is the dimension of the optimization space. Each X_h shows a possible response. And each possible response corresponds to a function $f(X_h)$ related to the optimization cost function.

Step 2: Ranking and grouping H frogs are arranged in descending rank based on performance of cost function. Position $P_x = [P_{x1}, P_{x2}, \dots, P_{xd}, \dots, P_{xD}]$, of the best frog based on cost function output in the population is separated. The population is divided into a memplexes, and there are c frogs in each memplex is defined as:

$$M_{o_1} = \{X_{o_1+\alpha(o_2-1)} \in Population \mid 1 \leq o_2 \leq c\} \tag{21}$$

$(1 \leq o_1 \leq \alpha)$

Step 3: Local search Inside each memplex, the nearby optimization handle is repeated for the desired number of iterations.

Step 3-1 Positions of the frogs in the memplex model, the best and the worst, are specified as $P_b = [P_{b1}, P_{b2}, \dots, P_{bd}, \dots, P_{bD}]$ and $P_w = [P_{w1}, P_{w2}, \dots, P_{wd}, \dots, P_{wD}]$, respectively. In this definition, P_w is updated based on:

$$D_{s_d} = \begin{cases} \min[INT(r \times (P_{bd} - P_{wd})), D_d^{max}] & P_{bd} - P_{wd} \geq 0 \\ \min[INT(r \times (P_{bd} - P_{wd})), -D_d^{max}] & P_{bd} - P_{wd} < 0 \end{cases}$$

$d = 1, 2, \dots, D$ (22)

$$P'_{wd} = P_{wd} + D_{s_d} \tag{23}$$

$$P'_{wd} = \begin{cases} Z_d^{max} & P'_{wd} > Z_d^{max} \\ Z_d^{min} & P'_{wd} \leq Z_d^{min} \end{cases}$$

(24)

where r is the random value in $[0,1]$ interval, D_{s_d} is the neighbor of the d_{th} decision variable, and D_d^{max} is the maximum neighbor of the d_{th} decision variable. Also, P'_{wd} is the updated position of the d_{th} decision variable.

Step 3-2 If the performance value of $P'_{wd} = [P'_{w1}, \dots, P'_{wd}, \dots, P'_{wD}]$ is better than P_w , then $P_w = P'_{wd}$; otherwise, P_b is defined as Eq. (22) and is replaced with P_x , and the position updating is performed repeatedly.

Step 3-3 If the cost function value of P_w is still better than P'_{wd} , then P_w is substituted by a random frog position.

Step 4: Shuffling and Global Search After a local search step, all memplexes values are mixed to form an updated population. Frogs are arranged and the optimal frog P_x is specified. After this level, the next grouping and local search results are performed until the determined number of global iterations is completed.

3 Experimental results

Using mammography images from the Mini-MIAS database [61], evaluation criteria were analyzed. The images downloaded from this database were scanned with *LJPEG* format in form of three-channel image with a size of 50 microns. The image resolution is 200 μm. Also, the downloaded images from Mini-MIAS database have a depth of 8 bits, and are in 1024 × 1024 dimensions. The Mini-MIAS mammograms have three channels. Therefore, due to the nature of the mammography imaging device, the images have been recorded in the gray-level form. Because the images are high-dimensional, we have resized them into 256 × 256 dimensions to reduce the computational complexity. The first column of data shows the reference for each image while the second column shows the background texture of the image. In the third column of the data, there are seven different classes of classified data as follows:

1. *CALC* Calcification
2. *CIRC* Well-defined/circumscribed masses
3. *SPIC* Spiculated masses
4. *MISC* Other, ill-defined masses
5. *ARCH* Architectural distortion
6. *ASYM* Asymmetry
7. *NORM* Normal.

The other column of data includes the severity of abnormal mass that comprised the letters *B* and *M* which are the abbreviation of benign and malignant, respectively. To evaluate the masses, ANFIS classifier output classes is employed based on 7 listed classes. By combining and integrating the solutions presented in Matlab

Table 2 Level 8 Pseudo-Zernike moments

Index	m	n	Dimen- sional- ity	Order	Zernike moments
1	0	0	1	0	PZM _{0,0}
2	1	1	2	1	PZM _{1,1}
3	2	0, 2	4	2	PZM _{2,2} , PZM _{2,0}
4	3	1, 3	6	3	PZM _{3,3} , PZM _{3,1}
5	4	0, 2, 2	9	4	PZM _{4,4} , PZM _{4,2} , PZM _{4,0}
6	5	1, 3, 5	12	5	PZM _{5,5} , PZM _{5,3} , PZM _{5,1}
7	6	0, 2, 4, 6	16	6	PZM _{6,6} , PZM _{6,4} , PZM _{6,2} , PZM _{6,0}
8	7	1, 3, 5, 7	20	7	PZM _{7,7} , PZM _{7,5} , PZM _{7,3} , PZM _{7,1}

programming environment, the proposed algorithm is constructed in three experimental steps.

3.1 Setting

In the features extraction step, 59 features were extracted from the image containing the location and condition of the mass. In Harlic matrix ($C_{m \times n}$), the most important features included contrast, energy, entropy, variance or the sum of squares, correlation, etc. In Table 1, some of these features along with their describing relationships are shown.

In the first step of describing features for each scale, the combination matrix of that scale was formed by placing all sub-bands together. Thereafter, the Co-occurrence matrix of that scale was constructed with the parameters $d = 1$ (pixel resolution distance) and angles 0, 45, 90 and 135°.

In the second step, the simulation feature extraction of mammographic image was carried out using

Table 1 Some features extracted by the GLCM algorithm

Index	Physical meaning of features	Mathematical expression
1	Energy	$\sum_{i=1}^n \sum_{j=1}^m C_{ij}^2$
2	Entropy	$-\sum_{i=1}^n \sum_{j=1}^m C_{ij} \log C_{ij}$
3	Contrast	$\sum_{i=1}^n \sum_{j=1}^m i - j C_{ij}$
4	Inverse difference moment (IDM)	$\sum_{i=1}^n \sum_{j=1}^m (C_{ij} / i - j ^k), i \neq j, k = 1, 2$
5	Homogeneity	$\sum_{i=1}^n \sum_{j=1}^m (C_{ij} / (1 + i - j))$
6	Sum-mean	$1/2 \times \sum_{i=1}^n \sum_{j=1}^m (iC_{ij} + jC_{ij})$
7	Variance	$1/2 \times \sum_{i=1}^n \sum_{j=1}^m ((i - \mu)^2 C_{ij} + (j - \mu)^2 C_{ij})$
8	Correlation	$\sum_{ij} (i - \mu)(j - \mu)C_{ij} / \sqrt{\text{var}(i)\text{var}(j)}$
9	Maximum probability (MP)	$MP = \max_{ij} C_{ij}$
10	Cluster tendency	$\sum_{i=1}^n \sum_{j=1}^m (i + j - 2\mu)^k C_{ij}, k = 2, 3$
11	Texture probability of run length of 2	$\sum_{i=1}^n \sum_{j=1}^m ((C_i - C_{ij})^2 C_{ij} / C_i^2)$
12	Inertia	$\sum_{i=1}^n \sum_{j=1}^m (i - \mu)^2 C_{ij}$
13	Intensity	$\sum_{i=1}^n \sum_{j=1}^m ijC_{ij}$

Pseudo–Zernike moments and the blocks were divided so that:

1. Block feature sets are inscribed on the picture
2. Block feature sets are inscribed on one fourth of the picture (image is divided into four equal portions)
3. 3 Block feature sets are inscribed blocks on one third transverse image (the image is divided horizontally into three equal parts)
4. Block feature sets are inscribed on one third longitudinal image (the image in vertical direction is equally divided into three parts).

Similarly in Table 2, the level 8 Pseudo–Zernike moments can be observed. In discrete wavelet transform, each mammogram can be scaled up to 3, 4 or 5 levels. The number of sub-bands in each of the levels is different. For level three, the number of sub-bands is 18 i.e. 1 + 16 + 1 and for level 4; the number of sub-bands is 50 i.e. 1 + 16 + 32 + 1, which are related to levels 1, 2, 3 and 4, respectively. The coefficients produce by wavelet transform to each 180 degrees are repetitive. As a result, half-sufficient sub-band was assumed for levels 2 and 3. Therefore at this stage of the four-level simulation, 26 sub-bands (i.e. 1 + 8 + 16 + 1) of the wavelet coefficients are generated and each sub-band is a set of coefficients. For feature extraction from 26 available sub-bands, the average information within each sub-band and standard deviation was measured. Each of the measured parameters produces a small amount.

Data in relation to Hold-out methods for fitness function of SA methods were selected. Back Propagation

Neural networks with a number of 6 neurons in the input layer, 8 neurons in the hidden layer and 4 neurons in output layer were tried to select efficient feature subset. The best number of selected members of aggregated features was from 18 to 25 which revealed the lowest classification error. Hence the numbers of features were saved in 6 groups of 4 and selected features became the input configurations of ANFIS. In order to simulate ANFIS, the configuration of network shown in Fig. 4. Process of the improved ANFIS by Shuffled Frog Leaping Algorithm has been shown in Fig. 5 schematic.

3.2 Assessments

The results of the implementation of preprocessing section for four samples of mammography images are shown in Fig. 6. In a series of sample images, neighborhood radius was assumed to be eight and redundant parts were eliminated from image-sets. In the first row of the images, examples of removing unwanted elements can be seen. Selection of the eight-neighborhood occurred due to the entire image analysis and applies to all images. According to masks obtained from the research by Oliver et al. [55] and benchmark assessments presented in (25) and (26), the pixels belonging to three sections: Background, Breast and Pectoral can be mutually compared. Since the goal is the separation of breast area from the rest regions, therefore we have:

$$OL_{Br-Pe} = \frac{2|Br \cap Pe|}{|Br| + |Pe|} \tag{25}$$

Fig. 4 The proposed structure of ANFIS classifier

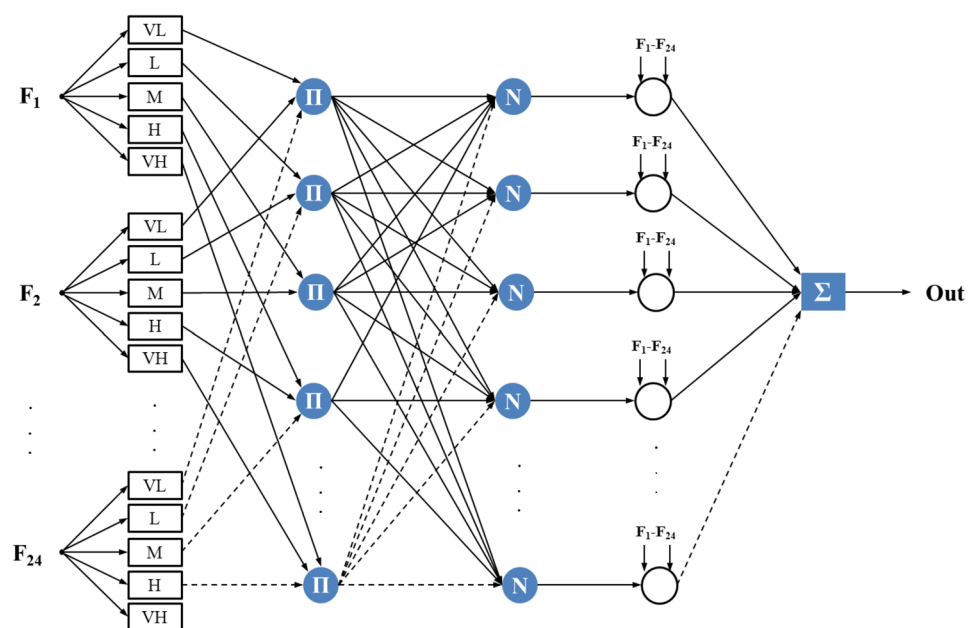
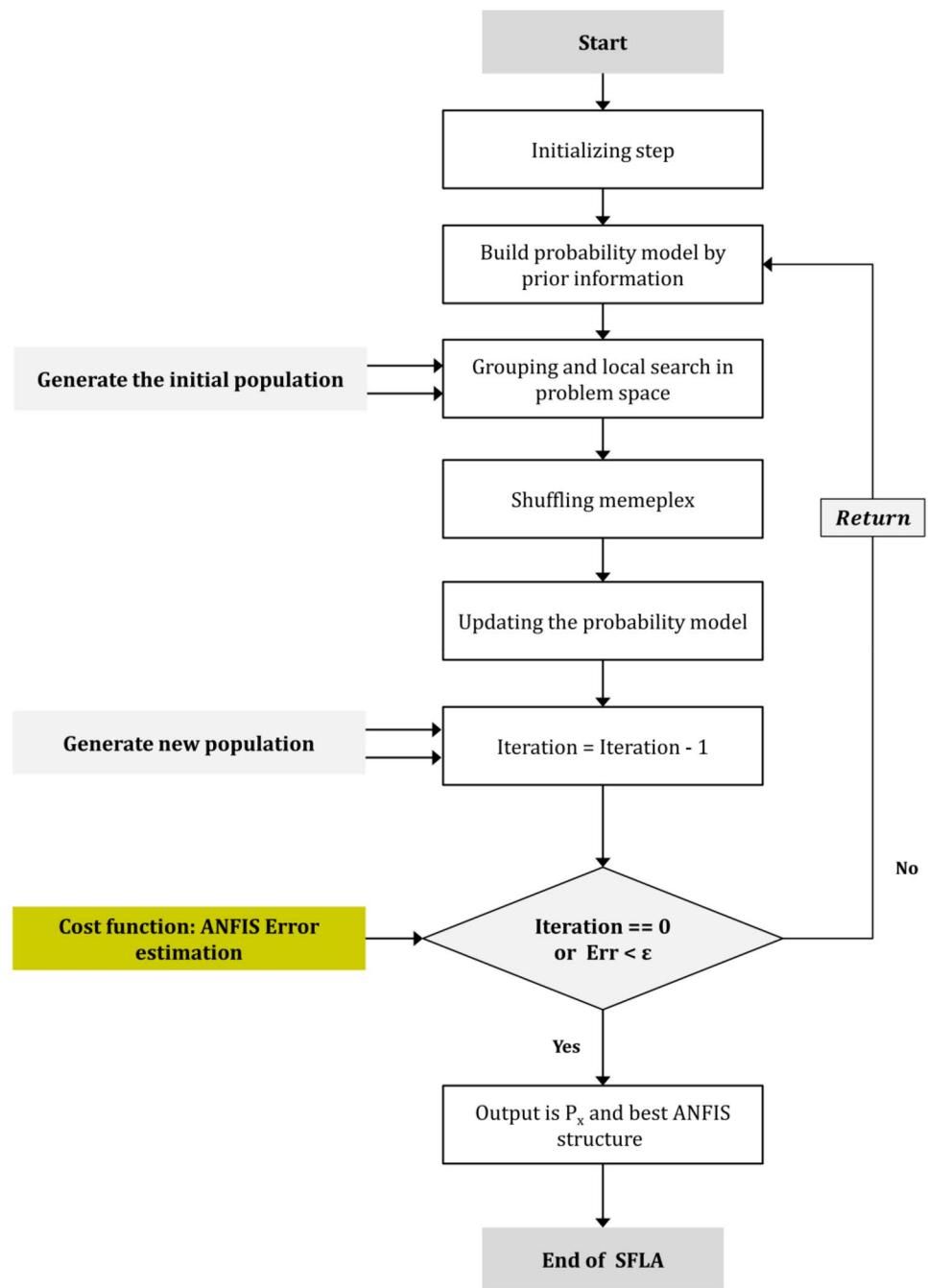


Fig. 5 Process of the improved ANFIS classifier by SFLA



$$OL_{Br-BG} = \frac{2|Br \cap BG|}{|Br| + |BG|} \tag{26}$$

Both equations represent the percentage of pixels overlapping for breast to pectoral area ($|Br \cap Pe|$) and breast to background ($|Br \cap BG|$). Table 3 shows the statistical results of segmentation in the pre-processing step for 322 image samples. In next level, with a choice of 4–6 spikes for mammography, mass location and its appearance can be segmented. Skilled radiologists were asked to identify the

location of the mass in the mammograms with different shapes and precisely map the locations of possible masses. They were blind to the database information, and even predicted the type of mass in images. There was a significant relationship between the stated prediction and the results of clustering at the end of the pre-processing step.

The data in the evaluation stage and proportional to K-fold method were divided into training and test data, and K-fold validation with $K=5$ were used. The train and test results are shown in Tables 4 and 5. Also, the training and test results in each table were provided and the

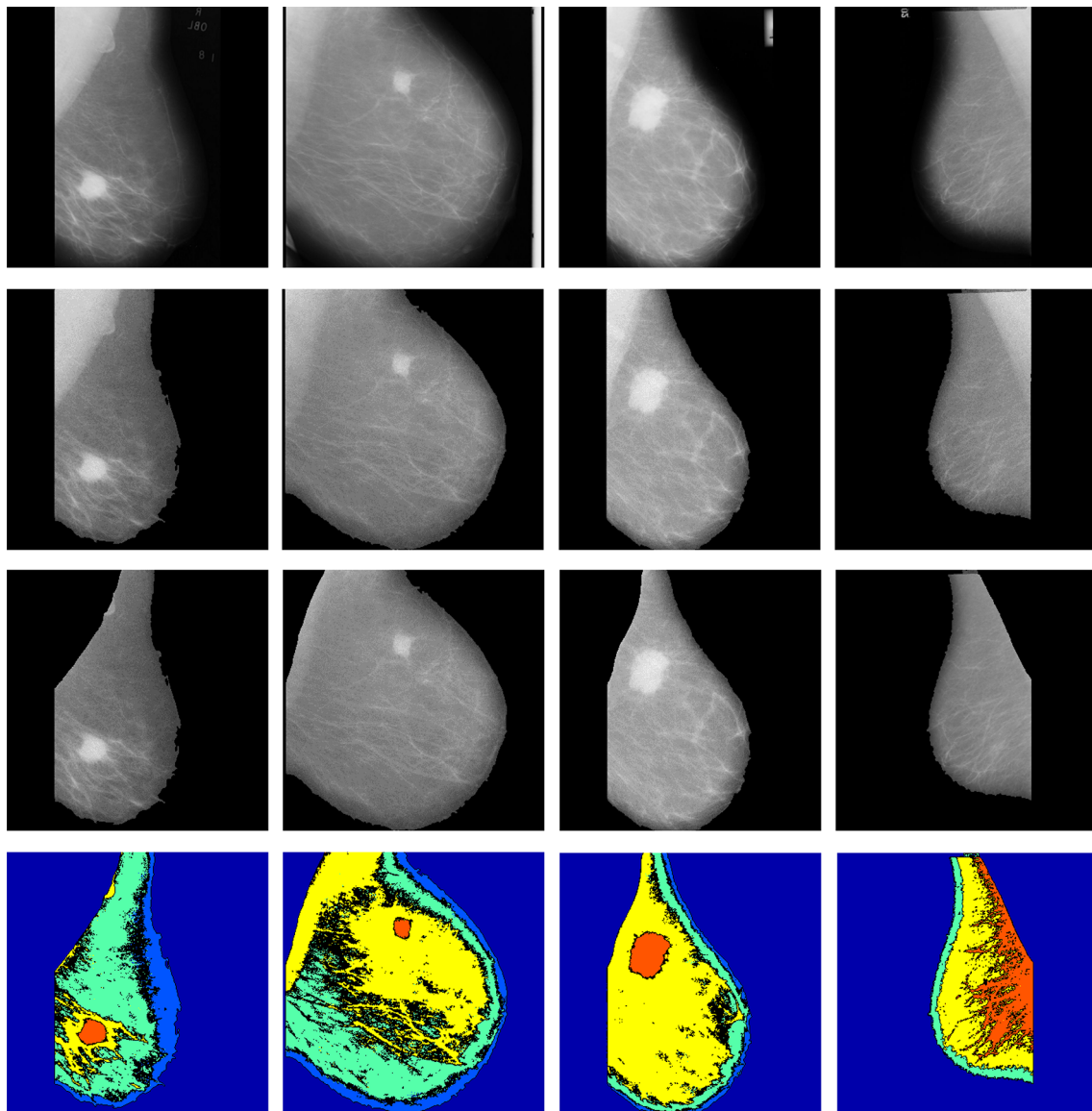


Fig. 6 This figure shows that the algorithm was implemented on 4 images. The first row: a section of an original mammography image. The second row: first step of the preprocessing of images

by removing unwanted objects. The third row: shows the result of pectoral segmentation while the fourth row: shows the results of region clustering by K-means algorithm

output was presented. In these tables, the accuracy has been computed based on confusion matrix for 7 classes of breast cancer. Also in Fig. 7, Receiver Operating Characteristic (ROC) curves were shown and calculation of the AUC shown in the images suggests the optimal performance of

the system in recognition of the different masses in mammography images.

For accurate comparison in the first curve, we use normal and abnormal images. We randomly split the data set into two parts (50% and 50%), with the 50% used to train the proposed algorithm and the 50% used as Hold-out cross validation to display ROC curve. As in several multi-class problems, the idea is to generally carry out pairwise comparison such as one class versus all other classes, and one class versus another class. On the other hand, we compared and plotted ROC curve for class 2 against classes 3, 4, etc. Thus in next step, we will compare and plot class 3 against classes 2, 4, etc.

Table 3 Shows the statistical results of segmentation in the preprocessing step

Step 1 and 2	OL _{Br-Pe}	OL _{Br-BG}	Mean
Before removing unwanted objects	< 23%	< 14%	< 18.5%
After removing unwanted objects	< 11%	< 6%	< 8.5%

Table 4 The accuracy results of applying the proposed algorithm by K-fold evaluation in train steps

	CALC	CIRC	SPIC	MISC	ARCH	ASYM	NORM	Accuracy (%)
K=1								
CALC	20	0	0	1	0	0	1	90.90
CIRC	0	18	1	0	0	0	1	90
SPIC	1	0	11	0	1	0	1	84.61
MISC	0	1	0	10	0	0	1	83.33
ARCH	1	0	0	0	11	1	1	78.57
ASYM	0	1	0	0	0	10	1	83.33
NORM	2	3	2	3	4	2	147	89.63
Mean accuracy								85.77
K=2								
CALC	19	1	0	1	0	0	1	86.36
CIRC	1	16	0	1	0	1	0	80
SPIC	0	0	12	0	0	1	1	85.71
MISC	0	0	0	10	1	0	1	83.33
ARCH	1	0	0	1	12	0	0	85.71
ASYM	0	1	1	0	1	9	0	75
NORM	1	2	1	3	2	4	151	92.07
Mean accuracy								84.02
K=3								
CALC	19	1	0	0	1	0	1	86.36
CIRC	0	18	1	0	0	0	1	90
SPIC	1	0	12	0	0	1	0	85.71
MISC	0	0	0	11	0	0	1	91.66
ARCH	0	0	0	0	12	1	1	85.71
ASYM	0	1	0	0	0	11	0	91.66
NORM	1	2	1	2	2	1	155	94.51
Mean accuracy								89.37
K=4								
CALC	18	1	0	1	1	0	1	81.81
CIRC	0	19	0	1	0	0	0	95
SPIC	1	0	12	0	0	1	0	85.71
MISC	0	0	0	11	0	0	1	91.66
ARCH	0	0	0	0	12	1	1	85.71
ASYM	0	1	0	0	0	11	0	91.66
NORM	1	0	1	2	2	1	157	95.73
Mean accuracy								89.61
K=5								
CALC	20	0	1	0	0	0	1	86.36
CIRC	1	18	0	1	0	0	0	85.71
SPIC	1	0	13	0	0	0	0	92.85
MISC	0	0	0	10	1	0	1	83.33
ARCH	1	0	0	0	12	0	1	85.71
ASYM	0	1	0	0	0	11	0	91.66
NORM	1	2	1	2	1	1	158	95.12
Mean accuracy								88.67
								87.52

4 Discussion

We have tested numerous clusters and calculated the results in the experiments separately. When the number

of clusters selected using the K-means was between 3 and 5, better outputs were obtained. This is shown in Fig. 8 by changing the number of clusters from 2–6 and calculating

Table 5 The accuracy results of applying the proposed algorithm by K-fold evaluation in test steps

	CALC	CIRC	SPIC	MISC	ARCH	ASYM	NORM	Accuracy (%)
K=1								
CALC	5	0	0	0	0	0	1	83.33
CIRC	1	4	0	0	0	0	0	80
SPIC	0	0	4	0	0	0	0	100
MISC	0	0	0	3	0	0	0	100
ARCH	0	0	0	0	3	0	1	75
ASYM	0	0	0	0	0	3	1	75
NORM	1	1	0	0	1	1	37	90.24
Mean accuracy								86.22
K=2								
CALC	5	0	0	0	0	0	1	83.33
CIRC	0	4	0	0	0	0	1	80
SPIC	0	0	3	0	0	0	1	75
MISC	0	0	0	3	0	0	0	100
ARCH	1	0	0	0	4	0	0	80
ASYM	0	1	0	0	0	3	0	75
NORM	2	1	0	1	0	1	36	87.80
Mean accuracy								83.01
K=3								
CALC	5	0	0	0	1	0	0	83.33
CIRC	0	4	0	0	0	0	1	80
SPIC	0	0	3	0	0	0	1	75
MISC	0	0	0	2	0	0	1	66.66
ARCH	0	0	0	0	4	0	0	100
ASYM	0	0	0	0	1	3	0	75
NORM	1	2	0	1	1	1	35	85.36
Mean accuracy								80.76
K=4								
CALC	5	0	0	0	0	0	1	83.33
CIRC	1	4	0	0	0	0	0	80
SPIC	0	0	3	0	0	0	1	75
MISC	0	0	0	2	0	0	1	66.6
ARCH	0	0	0	0	3	0	1	75
ASYM	0	0	1	0	0	3	0	75
NORM	1	2	2	0	1	2	33	84.83
Mean accuracy								77.10
K=5								
CALC	4	0	0	0	1	0	1	66.66
CIRC	0	5	0	0	0	0	0	100
SPIC	0	0	3	0	0	1	0	75
MISC	0	0	0	3	0	0	0	100
ARCH	0	0	0	0	3	0	1	75
ASYM	0	0	0	0	0	4	0	100
NORM	1	2	1	1	0	1	35	85.36
Mean accuracy								85.71
								82.56

the final classification accuracy and *AUC* to analyze the desired number of clusters.

By aggregating the features obtained from different descriptors a comparison has been made among their performance. All feature extraction procedures are shown

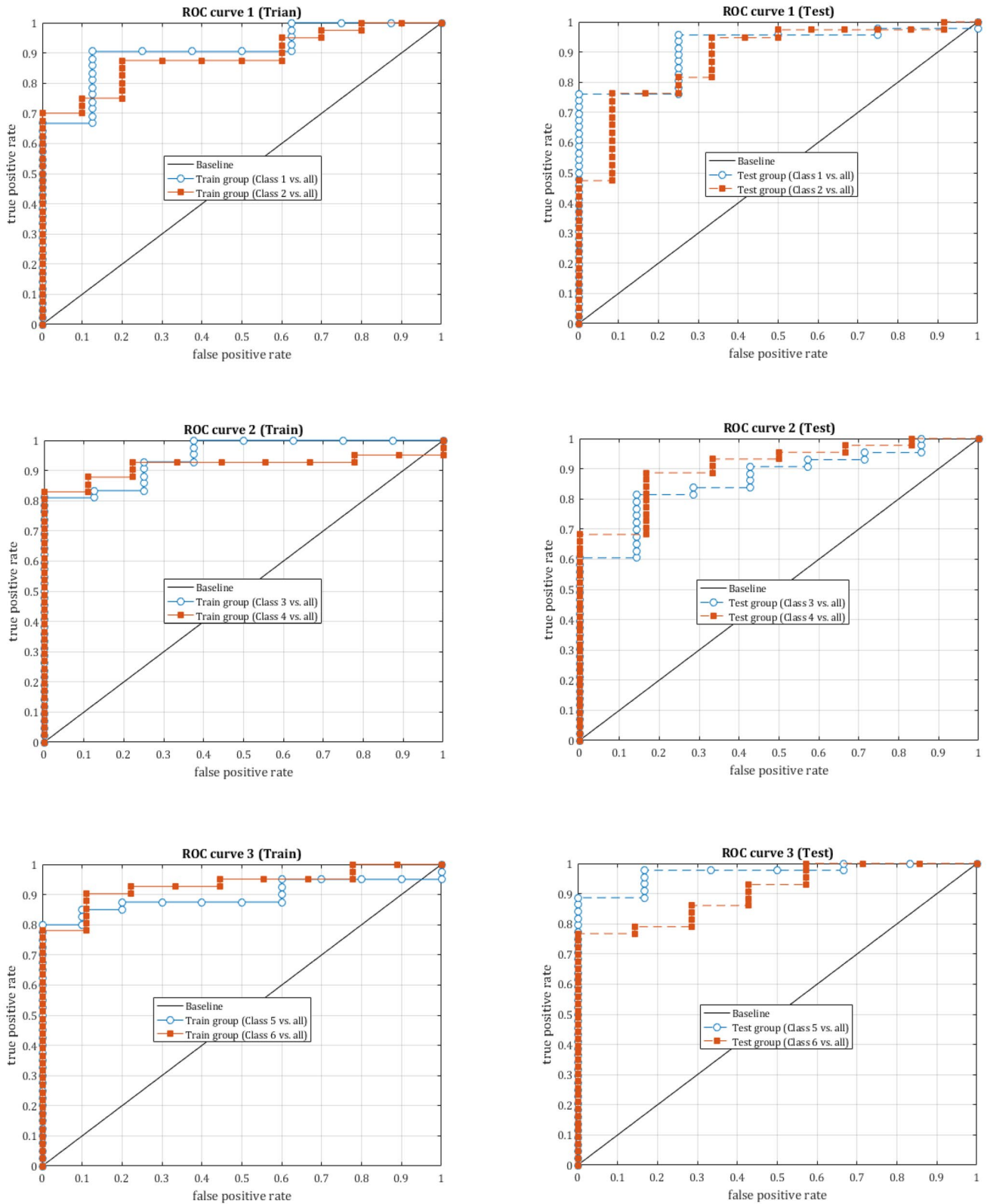


Fig. 7 ROC curves in the first plot showed one class versus other masses; while in this plots, random six ROC curves displayed one-against-all for various masses. The first column: ROC curves for random train groups. The second column: ROC curves for random test groups

in an in Fig. 9. Although the *GLCM*, *PZMs*, and *Wavelet* descriptors have allocated more suitable features, among the tissue features, to themselves, the feature aggregation has led to better results.

Due to using all data of the error matrix, the *Kappa* factor is used as the classification accuracy and fitness function assessments. This factor is defined as (26):

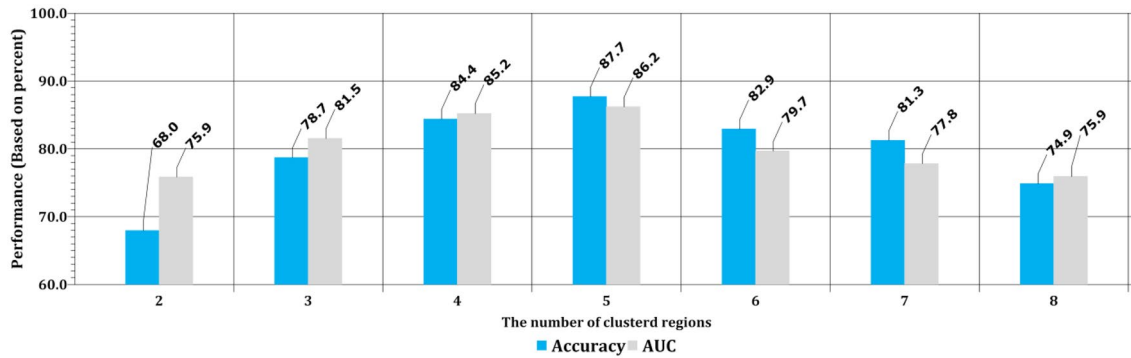


Fig. 8 Shows that the best value for the parameter K of K-means clustering

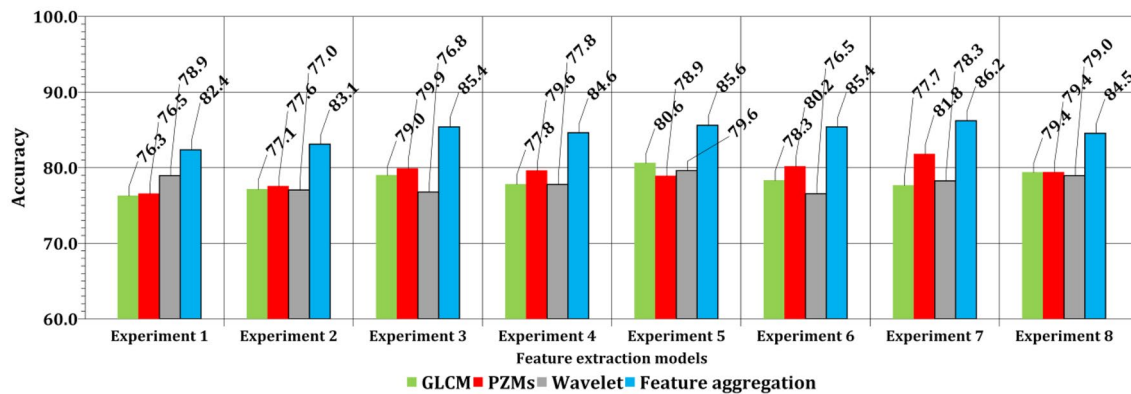


Fig. 9 Comparison between feature extraction schemes for mammograms images across multiple experiments

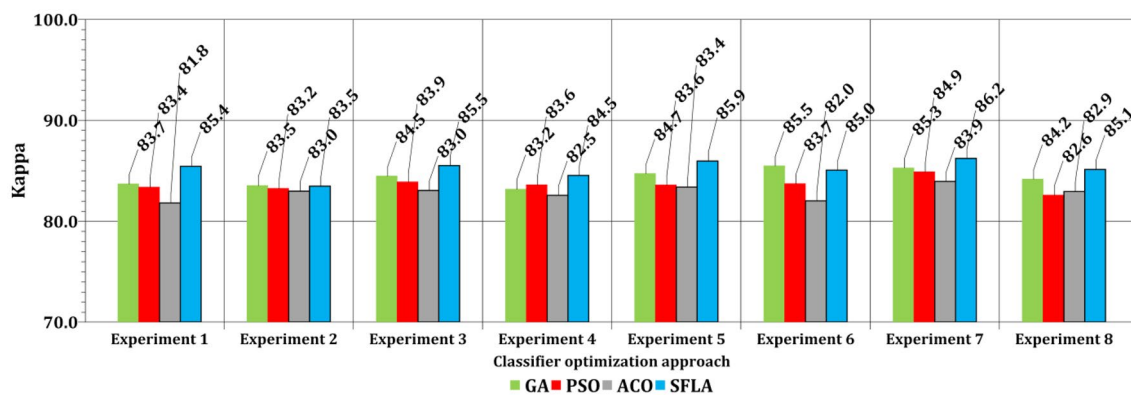


Fig. 10 Comparison between classifier optimization approaches for mammograms images across multiple experiments

$$Kappa = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})} \quad (27)$$

where N is the number of all data, r is the number of classes, x_{ij} denotes the elements on the main diagonal of the error matrix, x_{i+} is the marginal sum of rows, and x_{+j} shows the marginal sum of columns. Compared to different classification models based on Fig. 10, to optimize the ANFIS classifier with other methods such as *GA*, *PSO*, and *ACO*, the Kappa factor obtained from the *SFLA* is more acceptable. The performance of this algorithm in finding

global optimum is satisfactory and it can be used as an ANFIS optimizer algorithm in the classification of various masses of breast cancer.

The database had the specified label, but the images in the data were labeled as healthy or unhealthy and two radiologists were also asked to review the masses. Compared with the previous methods, total precision is at an appropriate level (Table 6). By calculating the AUC and sensitivity, the numbers equal to 94.14% for total masses and 96.89% for benign and malignant states were obtained, respectively. Because when the model is not optimized, the accuracy of applying the training data is higher than

Table 6 Comparison among proposed method and other breast cancer classification techniques

Methods	Descriptor	Classifier	Database	Results
Oliver et al. [34], two masses	Local binary pattern (LBP)	SVM	DDSM	80% area under curve
Jasmine et al. [51], micro-calcification	2-D wavelet transform	–	MIAS	87% accuracy
Rabottino et al. [23], two masses	Shape and texture descriptors	Fuzzy	DDSM	88% sensitivity
Mencattini et al. [24], two masses	Shape descriptor	–	DDSM	82% sensitivity
Rizzi et al. [25], micro-calcification	Adopting a wavelet decomposition	–	MIAS	98% sensitivity
Alofe et al. [26], two masses	2-D wavelet transform and texture descriptors	KNN	DDSM	71.93% sensitivity
Mazurowski et al. [37], two masses	Histogram features	–	DDSM and DBT	80% sensitivity
Zhang et al. [27], two masses	Shape descriptor	KNN	DDSM and MIAS	90% accuracy
Xu et al. [28], two masses	Shape and texture descriptors	–	DDSM	90% sensitivity
Olivier et al. [29], micro-calcification	Texture descriptor	SVM	DDSM	90% area under curve
Deserno et al. [38], two masses	Principle component analysis (PCA)	SVM	IRMA	80% Accuracy
Tamil et al. [35], two masses	Gabor descriptor	RBFNN	MIAS	85% sensitivity
Subashini et al. [30], two masses	Statistical descriptor	SVM	MIAS	94% accuracy
Kabbadj et al. [31], micro-calcification	Statistical and morphological descriptors	SVM	MIAS	99.60% sensitivity
Shanthi et al. [32], two masses	Statistical descriptor	SOM	MIAS	90.50% accuracy
Cabrera [33], two masses	Texture descriptor	–	DDSM	93% accuracy
Tahmasbi et al. [39], two masses	Zernike moments	MLP	MIAS	97.6% area under curve
Sharma et al. [40], two masses	Zernike moments	SVM	DDSM and IRMA	97.5% accuracy
Laroussi et al. [41], two masses	Zernike moments	MLP	DDSM	96% area under curve
Beura et al. [42], three masses	GLCM and 2-D wavelet transform descriptors	BPNN	MIAS and DDSM	94–98% accuracy
Singh et al. [22], two masses	Zernike moments	DE-WNN	MIAS	89% accuracy and 93.5% area under curve
Torrents et al. [36], two masses	Gabor descriptor	SVM	MIAS	88.4% accuracy
Fernandes et al. [50], two masses	Texture descriptor	ANFIS	MIAS	99% accuracy
Proposed Computer-Aided Algorithm	Feature extraction based on GLCM descriptor, Zernike moments and 2-D wavelet transform—Feature subset selection based on simulated annealing	MM-ANFIS (SFLA-ANFIS)	MIAS	Two masses; 97.5% accuracy and 95% area under curve Three masses; 93% accuracy and 91% area under curve Six masses; 82% accuracy (test) and 85% area under curve

that of the test data. The main reason for this event is the over-fitting problem and to prevent this challenge the model is tuned based on the SFLA algorithm, which helped optimize the accuracy of the test step. Thus, the results of the train and test steps for Mini-MIAS data are represented separately to show that the over-fitting challenge for a large number of classes has been considered.

It can be seen that the performance of the algorithm for identifying benign and malignant masses as well as their separation compared to methods such as [22, 30–34, 36, 39–42] is effective. Although the total precision compared to methods such as [30, 31, 50] is less, it should be noted that if the algorithm to be implemented in two stages on two categories of data is in line with the mentioned procedures, then the binary classification accuracy and AUC (i.e. healthy or unhealthy), will be higher than 98.6%. Thus, by diagnosing healthy individuals from patients, this algorithm offers a better performance from priory algorithms [30, 31, 33, 39–42, 50]. By distinguishing patients from healthy people, the statistical population is limited to 121 images; among them, circumscribed masses comprised 25 images, Spiculated masses with 19 images, ill-defined masses with 15 images, Architectural distortion with 19 images, Asymmetry with 15 images and Calcification with 28 images.

We applied the algorithm again to the class of diseases and good precision (above 92%) was obtained for the six different classes. Although such methods [22, 32, 33, 36, 39–42, 50] have a favorable level of accuracy and sensitivity, the lack of integrity in discrimination and segmentation of all masses was criticized. Therefore, the first difference and a key advantage of the presented solution are in recognition of the different kind of masses.

In identifying micro-calcification, methods [29, 31, 51], respectively have functions equal to 87, 90 and 99.60%, but in their study, the F_1 score does not result from sensitivity and specificity and the time taken for the procedure [31] is unclear, because this method made use of two categories of features. Unlike the former methods [22, 32, 33, 36, 39–42, 50], which did not assess data conclusiveness and the correlation of algorithm performance by the radiologist opinion, in this study, the p value in the proposed algorithm showed a significant correlation between the output of the proposed algorithm and the Radiologist opinion ($p < 0.05$). The comparison of the results with radiologists and other similar methods is a proper reason to reject the Null hypothesis (H_0). Since the test result was not placed in the acceptable area H_0 , H_0 is not accepted ($\alpha = 0.05$ and thus $-Z_{\alpha-1} = -1.65$). This means that Confidence Interval is more than 95%, and despite the large number of mammography images and the evaluation of K-fold, outputs are closer to reality.

In addition, the strengths of the algorithm can be seen as a recognition tool. In some studies, tumor location based segmentation is in line with the recognition, while with the correct segmentation techniques; isolation of the breast and pectoral as well as ROI in the pre-processing step was performed. In few researches, heuristic algorithm is used to reduce the dimensions and choose the best subset of features. Also, the accuracy of the algorithm showed its ability to histological features and statistical analysis of mammography images. In some other studies, the aggregate descriptors were used for the recognition of the best features; for instance, GLCM and wavelet feature was used by Beura et al. [42], while the number of statistical descriptors and texture in the current context is three extraction features. Furthermore, ANFIS classifier, adaptive model of fuzzy inference system and neural network are special abilities for the classification of multiple classes [39, 48, 62, 63].

5 Conclusion

The recognition and early detection of breast cancer in women can be a strategy for fighting this disease. Thus, the need for an efficient system with the ability to automatically separate different classes of mass labels is necessary. In this paper, a combination method including feature extraction of mammography image based on aggregating various characteristics of three texture and statistical descriptors and also selection of efficient features by the algorithm SA was offered in the first place. By separating the target region in mammography image, the memetic meta-heuristic adaptive neuro-based fuzzy inference system (MM-ANFIS) classifier for classification was employed to verify the attained classification of above 90%. Although the different algorithms for estimating the presence or absence of this disease were proposed, the form and shape of the suspected masses can be effective in combating and preventing this disease in early stage. In future, the authors plan to optimize the feature extraction pattern, feature selection and modification of classifiers adaptive features to reduced feature dimensions and at the same time, reducing data processing time increases the precision.

Acknowledgements We gratefully acknowledge the generous support of Meybod University and Dr. Arnau Oliver for this radiology research, as well as their mammography masks, without which, this study could not have been accomplished.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Breast Cancer* 136(5):359–386
- Siegel RL, Jemal A, Wender RC, Gansler T, Ma J, Brawley OW (2018) An assessment of progress in cancer control. *CA Cancer J Clin* 68:329–339
- Siegel RL, Miller KD, Jemal A (2019) Cancer statistics. *CA Cancer J Clin* 69:7–34
- Aldridge RW, Nellums LB, Bartlett S, Barr AL, Patel P, Burns R et al (2018) Global patterns of mortality in international migrants: a systematic review and meta-analysis. *Lancet* 392:2553–2566
- Dheeba J, Singh NA, Selvi ST (2014) Computer-aided detection of breast cancer on mammograms: a swarm intelligence optimized wavelet neural network approach. *J Biomed Inform* 49:45–52
- Colditz GA, Bohlke K (2014) Priorities for the primary prevention of breast cancer. *CA Cancer J Clin* 64(3):186–194
- DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, Alteri R, Robbins AS, Jemal A (2014) Cancer treatment and survivorship statistics. *CA Cancer J Clin* 64(4):252–271
- DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A (2015) International variation in female breast cancer incidence and mortality rates. *Cancer Epidemiol Biomark Prev* 24(10):1495–1506
- Martini N, Koukou V, Sotiropoulou P, Michail C, Kandarakis I, Nikiforidis G, Fountos G (2014) A novel non-invasive method substituting breast cancer biopsies. *Physica Med* 30:84
- Olsen AH, Lynge E, Njor SH, Kumle M, Waaseth M, Braaten T, Lund E (2013) Breast cancer mortality in Norway after the introduction of mammography screening. *Int J Breast Cancer* 132(1):208–214
- Gøtzsche PC, Jørgensen KJ (2013) Screening for breast cancer with mammography. The Cochrane Library, London
- Welch HG, Passow HJ (2014) Quantifying the benefits and harms of screening mammography. *JAMA Int Med* 174(3):448–454
- Taylor PM, Champness J, Given-Wilson RM, Potts HWW, Johnston K (2004) An evaluation of the impact of computer-based prompts on screen readers' interpretation of mammograms. *Br J Radiol* 77(913):21–27
- Zheng B, Yoon SW, Lam SS (2014) Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Syst Appl* 41(4):1476–1482
- Chandrashekar G, Sahin F (2014) A survey on feature selection methods. *Comput Electr Eng* 40(1):16–28
- Tan M, Qian W, Pu J, Liu H, Zheng B (2015) A new approach to develop computer-aided detection schemes of digital mammograms. *Phys Med Biol* 60(11):4413
- Jalalian A, Mashohor SB, Mahmud HR, Saripan MIB, Ramli ARB, Karasfi B (2013) Computer-aided detection-diagnosis of breast cancer in mammography and ultrasound: a review. *Clin Imaging* 37(3):420–426
- Tang J et al (2009) Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Trans Inf Technol Biomed* 13(2):236–251
- Lo CM, Chen RT, Chang YC, Yang YW, Hung MJ, Huang CS, Chang RF (2014) Multi-dimensional tumor detection in automated whole breast ultrasound using topographic watershed. *IEEE Trans Med Imaging* 33(7):1503–1511
- Rezaee K, Haddadnia J (2013) Designing an algorithm for cancerous tissue segmentation using adaptive k-means clustering and discrete wavelet transform. *J Biomed Phys Eng* 3:93
- Hadadnia J, Rezaee K (2013) Extraction and 3D segmentation of tumors-based unsupervised clustering techniques in medical images. *Iran J Med Phys* 10:95–108
- Singh SP, Urooj S (2016) An improved CAD system for breast cancer diagnosis based on generalized Pseudo-Zernike moment and Ada-DEWNN classifier. *J Med Syst* 40(4):1–13
- Rabottino G, Mencattini A, Salmeri M, Caselli F, Lojaco R (2008) Mass contour extraction in mammographic images for breast cancer identification. In: Proceedings of the 16th IMEKO TC4 symposium, exploring new frontiers of instrumentation and methods for electrical and electronic measurements, Florence, Italy
- Mencattini A, Rabottino G, Salmeri M, Salicone S (2009) Uncertainty propagation for the assessment of tumoral masses segmentation. In: Proceedings of the IEEE international workshop on advanced methods for uncertainty estimation in measurement (AMUEM)—IEEE, pp 39–43
- Rizzi M, D'Aloia M, Castagnolo B (2009) Computer aided detection of microcalcifications in digital mammograms adopting a wavelet decomposition. *Integr Comput Aided Eng* 16(2):91–103
- Aloffe MA, Youssef ABM, Kadah YM, Mohamed AS (2008) Development of a computer-aided classification system for cancer detection from digital mammograms. In: Proceedings of the radio science conference, pp 1–8
- Zhang M, Chai Y, Wang J (2011) An integrated method for breast mass segmentation in digitized mammograms. In: Proceeding of the 3rd international conference on advanced computer control (2011), pp 214–218
- Xu S, Liu H, Song E (2011) Marker-controlled watershed for lesion segmentation in mammograms. *J Digit Imaging* 24(5):754–763
- Oliver A, Lladó X, Freixenet J, Martí J (2007) False positive reduction in mammographic mass detection using local binary patterns. In: Proceeding in the international conference on medical image computing and computer-assisted intervention, pp 286–293
- Subashini TS, Ramalingam V, Palanivel S (2010) Automated assessment of breast tissue density in digital mammograms. *Comput Vis Image Underst* 114(1):33–43
- Kabbadj Y, Regragui F, Himmi MM (2012) Microcalcification detection using a fuzzy inference system and support vector machines. In: Proceeding of the IEEE international conference on multimedia computing and systems, pp 312–315
- Shanthi S, Bhaskaran VM (2012) Computer aided detection and classification of mammogram using self-adaptive resource allocation network classifier. In: Proceeding of the IEEE international conference on in pattern recognition, informatics and medical engineering, pp 284–289
- Guzmán-Cabrera R, Guzmán-Sepúlveda JR, Torres-Cisneros M, May-Arriola DA, Ruiz-Pinales J, Ibarra-Manzano OG, Aviña-Cervantes G, Parada AG (2013) Digital image processing technique for breast cancer detection. *Int J Thermophys* 34(8–9):1519–1531
- Oliver A, Torrent A, Lladó X, Tortajada M, Tortajada L, Sentis M, Freixenet J, Zwigelaar R (2012) Automatic microcalcification and cluster detection for digital and digitised mammograms. *Knowl-Based Syst* 28:68–75
- Dheeba J, Selvi ST (2010) Screening mammogram images for abnormalities using radial basis function neural network. In: Proceeding of the IEEE international conference on communication control and computing technologies, pp 554–559
- Torrents-Barrena J, Puig D, Melendez J, Valls A (2016) Computer-aided diagnosis of breast cancer via Gabor wavelet bank and binary-class SVM in mammographic images. *J Exp Theor Artif Intell* 28(1–2):295–311
- Mazurowski MA, Lo JY, Harrawood BP, Tourassi GD (2012) Mutual information-based template matching scheme for detection of

- breast masses: from mammography to digital breast tomosynthesis. *J Biomed Inform* 44(5):815–823
38. Deserno TM, Soiron M, de Oliveira JE, Araújo ADA (2012) Computer-aided diagnostics of screening mammography using content-based image retrieval. In: *Proceeding of the SPIE medical imaging*, pp 831527–831527
 39. Tahmasbi A, Saki F, Shokouhi SB (2011) Classification of benign and malignant masses based on Zernike moments. *Comput Biol Med* 41(8):726–735
 40. Sharma S, Khanna P (2015) Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM. *J Digit Imaging* 28(1):77–90
 41. Laroussi MG, Ayed NGB, Masmoudi AD, Masmoudi DS (2013) Diagnosis of masses in mammographic images based on Zernike moments and local binary attributes. In: *2013 IEEE World congress on computer and information technology (WCCIT)*, pp 1–6
 42. Beura S, Majhi B, Dash R (2015) Mammogram classification using two dimensional discrete wavelet transforms and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing* 154:1–14
 43. Raghavendra U, Acharya UR, Fujita H, Gudigar A, Tan JH, Chokkadi S (2016) Application of gabor wavelet and locality sensitive discriminant analysis for automated identification of breast cancer using digitized mammogram images. *Appl Soft Comput* 46:151–161
 44. Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, den Heeten A, Karssemeijer N (2017) Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal* 35:303–312
 45. Drukker K, Huynh BQ, Giger ML, Malkov S, Avila JI, Fan B, Joe B, Kerlikowske K, Drukteinis JS, Kazemi L, and Pereira MM (2017) Deep learning and three-compartment breast imaging in breast cancer diagnosis. In: *SPIE medical imaging, international society for optics and photonics* (2017), 101341F
 46. Khan S, Islam N, Jan Z, Din IU, Rodrigues JJC (2019) A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recogn Lett* 125:1–6
 47. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W (2019) Deep learning to improve breast cancer detection on screening mammography. *Sci Rep* 9:1–12
 48. Xie W, Li Y, Ma Y (2016) Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing* 173:930–941
 49. Muduli D, Ratnakar D, Banshidhar M (2020) Automated breast cancer detection in digital mammograms: a moth flame optimization based ELM approach. *Biomed Signal Process Control* 59:101912
 50. Fernandes FC, Brasil LM, Lamas JM, Guadagnin R (2010) Breast cancer image assessment using an adaptive network-based fuzzy inference system. *Pattern Recognit Image Anal* 20(2):192–200
 51. Jasmine JL, Govardhan A, Baskaran S (2009) Microcalcification detection in digital mammograms based on wavelet analysis and neural networks. In: *Proceedings of the 10th 2009 international conference on in control, automation, communication and energy conservation* (2009), pp 1–6
 52. Warren LM, Mackenzie A, Cooke J, Given-Wilson RM, Wallis MG, Chakraborty DP, Dance DR, Bosmans H, Young KC (2012) Effect of image quality on calcification detection in digital mammography. *Med Phys* 39(6):3202–3213
 53. Tzikopoulos SD, Mavroforakis ME, Georgiou HV, Dimitropoulos N, Theodoridis S (2011) A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. *Comput Methods Programs Biomed* 102(1):47–63
 54. Mustra M, Grgic M (2013) Robust automatic breast and pectoral muscle segmentation from scanned mammograms. *Sig Process* 93(10):2817–2827
 55. Oliver A, Lladó X, Torrent A, Martí J (2014) One-shot segmentation of breast, pectoral muscle, and background in digitised mammograms. In: *Proceeding of the IEEE international conference on image processing (ICIP)*, pp 912–916
 56. Lee S, Kim G, Kim S (2011) Self-adaptive and dynamic clustering for online anomaly detection. *Expert Syst Appl* 38(12):14891–14898
 57. Salvador S, Chan P (2004) Determining the number of clusters-segments in hierarchical clustering-segmentation algorithms. In: *Proceeding of the 16th IEEE international conference on tools with artificial intelligence*, pp 576–584
 58. Wang W, Mottershead JE, Mares C (2009) Mode-shape recognition and finite element model updating using the Zernike moment descriptor. *Mech Syst Signal Process* 23(7):2088–2112
 59. Cvetkovic D, Übeyli ED, Cosic I (2008) Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: a pilot study. *Digit Signal Proc* 18(5):861–874
 60. Rezaee A, Rezaee K, Haddadnia J, Taheri Gorji H (2020) Supervised meta-heuristic extreme learning machine for multiple sclerosis detection based on multiple feature descriptors in MR images. *SN Appl Sci* 2:1–19
 61. Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, Ricketts I, Stamatakis E, Cerneaz N, Kok S, Taylor P (1994) The mammographic image analysis society digital mammogram database. In *Exerpta Medica. Int Congr Ser* 1069:375–378
 62. Abraham A (2005) Adaptation of fuzzy inference system using neural learning. In: *Proceeding of the Fuzzy systems engineering*, Springer, Berlin, pp 53–83
 63. Tahmasebi P, Hezarkhani A (2012) A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. *Comput Geosci* 42:18–27

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.