Journal of King Saud University – Computer and Information Sciences xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Classification of masses in digital mammograms using Biogeography-based optimization technique

Shankar Thawkar [a,*], Ranjana Ingolikar [b]

[a] Department of Electronics and Computer Science, R. T. M. Nagpur University, Nagpur, Maharashtra, India
[b] Department of Computer Science, S.F.S. College, Seminary Hills, Nagpur, Maharashtra, India

ARTICLE INFO

Article history:
Received 21 November 2017
Revised 27 December 2017
Accepted 15 January 2018
Available online xxxx

Keywords:
Biogeography-based optimization
Artificial neural network
Adaptive neuro-fuzzy inference system
Digital mammography
ROC curve
Classification

ABSTRACT

Breast cancer is a leading cause of death in women in both developed and developing countries. Design and development of computer-based systems can assist radiologists in the effective treatment of breast cancer. For the design of an efficient classification system, efficient feature selection techniques must be used to reduce complexity of feature space in digital mammogram classification. The proposed methodology aims to explore use of Biogeography-based optimization to select a subset of features. Adaptive neuro-fuzzy inference system and artificial neural network are employed to evaluate fitness of the selected features. The features selected are used to train and test adaptive neuro-fuzzy inference system and artificial neural network classifiers. The experiment employed over 651 mammograms. The classification results shows that Biogeography-based optimization with adaptive neuro-fuzzy inference system is superior to Biogeography-based optimization with artificial neural network. Adaptive neuro-fuzzy inference system classifier achieve an accuracy of 98.92% with sensitivity of 99.10%, specificity of 98.72% and area under curve $A_Z = 0.999 \pm 0.000$. Outcomes achieved with the proposed Biogeography-based optimization with adaptive neuro-fuzzy inference system are far better as compared to some recent work.

1. Introduction

Breast cancer is a leading cause of death in women in both developed and developing countries. Currently, there is no method available for prevention of breast cancer. Mammography is the best available tool for detection of breast cancer in the primary stage (DeSantis et al., 2011). A mass recognized by mammography is one of the important symptoms of breast cancer. Radiologists diagnose such masses by reading mammograms, which is not an easy task. Therefore, suspicious tissues are removed from the breast to check for the presence of cancer using breast biopsy. Available facts indicate that more than 60–70% biopsies of suspicious masses turn out to be benign cases. The use of a computer based diagnosis system can help to minimize unnecessary biopsies. Such systems can act as a second opinion for the radiologist for effective diagnosis of breast cancer and help to minimize mortality rate. The use of image processing and machine learning algorithms for detection and classification of masses in digital mammograms would be an easier method, but still it is a challenging area of study. This article focuses on the design of an efficient technique for feature selection and classification of masses in digital mammograms. Features submitted to the classifier without feature selection would affect the accuracy of the classifier; this is why feature selection should be applied before classification. The main objective of the feature selection process is to remove irrelevant or redundant features, to improve accuracy of the classifier and reduce computational burden.

2. Related work

Feature selection and classification of suspicious breast masses pose the most challenging research area. In recent years, several studies have been conducted in the field of mammography. A brief overview of Biogeography-based optimization and adaptive neuro-fuzzy inference system used in some recent applications are presented below.

* Corresponding author.
E-mail address: Shankar.thawkar@gmail.com (S. Thawkar).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2018.01.004
1319-1578/

Please cite this article in press as: Thawkar, S., Ingolikar, R. Classification of masses in digital mammograms using Biogeography-based optimization technique. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.01.004

Study of how natural biogeography is used to solve real world problems is presented in Simon (2008). The study comes up with a new field, Biogeography-based optimization (BBO), with practical uses to solve optimization problems. BBO has several features in common with biology-based techniques.

A model with wavelet analysis and fuzzy-neural approach for the detection of abnormality in mammograms is proposed in Mousa et al. (2005).

Chang and Chang (2006) proposed an Adaptive neuro-fuzzy inference system to predict water level in reservoirs. For demonstrating the capabilities of ANFIS, two models were developed. One uses the human decision and reservoir data set, and the other considers the data set without human decision. The results show that ANFIS model with human decision, as input is superior.

Lahmiri and Boukadoum (2011a) proposed a method based on discrete wavelet transform and Gabor filter for feature extraction from mammogram images. The method uses discrete wavelet transform and Gabor filter. The features namely average and standard deviation obtained from Gabor-filtered images are used to train and test support vector machine (SVM) classifier with polynomial kernel. The method was tested on 100 mammogram images and achieves an accuracy of 98%.

A new methodology based on discrete wavelet transform (DWT) and Radon transform for mammogram classification was proposed in Lahmiri and Boukadoum (2011b). The high-high (HH) sub-band mammogram image is obtained by DWT. Features namely energy and entropy obtained from Radon transform signals are used to train and test support vector machine (SVM) classifier with polynomial kernel. The method was tested on 100 mammogram images and achieves an accuracy of 97% (±0.031). Method achieves high accuracy with less computation time and requires minimum parameters to be tune

A PSO based artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) is proposed for classification of masses in Huang et al. (2012).

Lahmiri and Boukadoum (2013) proposed a method for automatic feature extraction and classification of biomedical images. The method uses two-dimensional discrete wavelet transform and Gabor filter bank. Features namely entropy and uniformity obtained from Gabor-filtered images are used to train and test support vector machine (SVM) classifier. The method achieves an accuracy of 97.36% ±0.02 with little computation time.

Zhang et al. (2016a) proposed a model with Biogeography-based optimization and feedforward neural network (FNN) for the classification of fruit images. The method uses the principle component analysis (PCA) technique for feature reduction. Results show that BBO-FNN outperforms five state of the art techniques.

Zhang et al. (2016b) proposed Multilayer Perceptron (MLP) for prediction of abnormal breasts in digital mammograms and used Biogeography-based optimization (BBO) to train the multilayer perceptron model. The method uses fractional Fourier entropy method to extract global features and Welch's t-test for feature selection.

Raghavendra et al. (2016) proposed method for automatic classification of mammogram images into normal, benign and malignant classes. The method uses Gabor wavelet for feature extraction and obtained reduce feature set by Locality Sensitive Discriminant Analysis. Several classifiers are used for the classification of mammograms; they are Linear Discriminant Analysis and Quadratic Discriminant Analysis, k-Nearest Neighbor, Naïve Bayes Classifier, Probabilistic Neural Network, Support Vector Machine, AdaBoost and Fuzzy Sugeno. The performance of all the classifiers is evaluated using 690 mammograms. The method achieves an accuracy of 98.69% for K-NN classifier with 10-fold cross validation. The method helps to improve breast cancer diagnosis.

A hybrid method for classification of clusters of microcalcifications (MCCs) is proposed in Khehra and Pharwaha (2017). They use genetic algorithm (GA), particle swarm optimization (PSO) and BBO methods for selection of optimal features based on the classifier's accuracy. Support vector machine issued as a classifier. Result shows that BBO-based feature selection is superior to GA-based and PSO-based methods.

Acharya et al. (2017) proposed a method for the characterization of cardiac abnormalities. The method uses Continuous wavelet transform with contourlet and shearlet transform for obtaining the features. To select optimal features, improved binary particle swarm optimization method is used. These selected features are used to characterize cardiac abnormalities using decision tree and K-nearest neighbor classifiers. The proposed method achieves an accuracy of 99.55% for contourlet transform and 99.01% using shearlet transform.

Raghavendra et al. (2018) proposed a novel computer aided diagnosis system for the automated detection of coronary artery disease. An echocardiography images are decomposed into subband images using double density-dual tree discrete wavelet transform (DD-DTDWT). The method uses marginal fisher analysis (MFA) and feature ranking method for the selection of optimal features. The method achieves an accuracy of 96.05% for linear discriminant classifier.

While reviewing literature we observed that Biogeography based optimization has been successfully applied in many applications for feature selection and classification. The encouraging results of BBO in other applications motivate us to implement it for feature selection and classification of masses in digital mammograms.

The rest of the paper organized as: Proposed framework, presented in Section 3. Section 4 presents results of different methods and discussion about their performances. Computational Complexity of the proposed method is given in Section 5. Section 6 presents conclusion of the paper.

3. Proposed framework

The proposed framework for classification of masses in digital mammograms consists of two main steps. In the first, we used Biogeography based optimization for feature selection, and to select optimal features, adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN) were used. In the second step, optimal features were selected by BBO used to classify suspicious masses into benign and malignant using, ANFIS and ANN. Overview of proposed framework is as shown in Fig. 1.

3.1. Feature selection

Features extracted from each of the segmented masses were categorized into three types: intensity based, texture based and shape based. Six features based on intensity, eleven features based on texture and eight features based on shape were extracted from each of the suspicious masses as shown in Table 1 (Thawkar and Ingolikar, 2017a,b). In Table 1, Entropy1 denotes Entropy computed from histogram analysis (Intensity based features) while Entropy computed from Gray Level Co-occurrence Matrix (Texture based features) is denoted by Entropy2.

Optimal feature selection is the most significant step of any classification system; any feature set containing a large number of features may hamper the performance of the classifier (Thawkar and Ingolikar, 2017a). Therefore, it is important to select only features that would improve classifier accuracy by removing unnecessary features (Sameti et al., 1997; Li et al., 2001).

The proposed classification method based on Biogeography-based optimization, discussed in the following section.

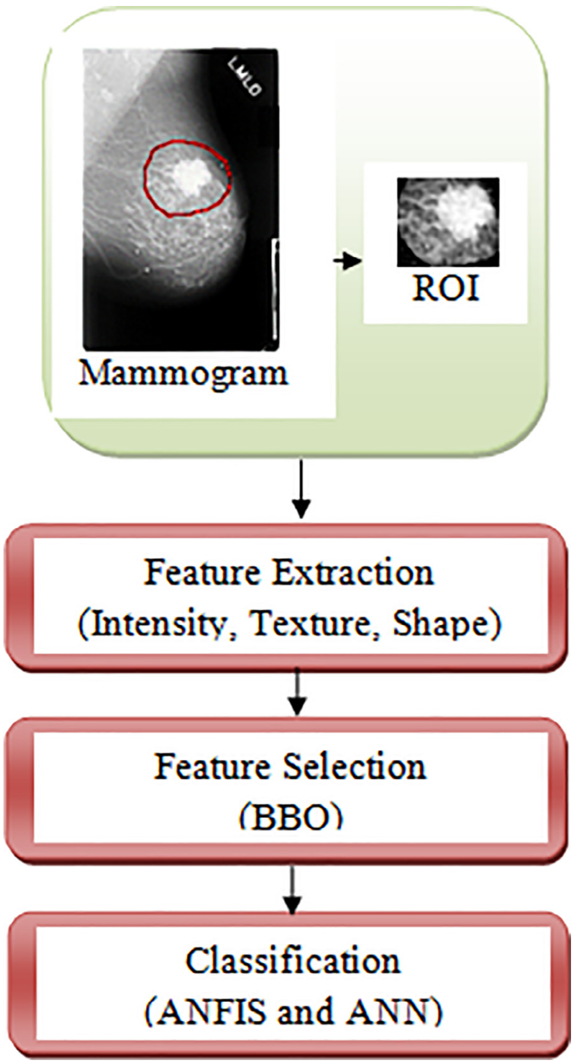

Fig. 1. Proposed framework.

3.1.1. Biogeography

The mathematical model about biogeography developed by Robert MacArthur and Edward Wilson in 1960 and published their work as "*The theory of Island Biogeography*" in 1967 (MacArthur and Wilson, 1967). The model describe migration of species from one island to another, arrival of species to island and extinct of species. Island is a "habitat" which is geographically separated from other habitats by water (Simon, 2008). In biogeography, the goodness of the living conditions in habitats are express by dependent variables called habitat suitability index, denoted by HSI. The value of HSI is high if the habitat is highly favorable for species as residence, and low if it is not suitable as residence for species (Wesche et al., 1987). The features associated with HSI are rainfall, variety of vegetation, temperature, ground area and topographic features. The habitability characterized by independent variables is called suitability index variable (SIV). The high HSI value of habitat tends to have more number of species while low HSI value indicates less number of species in the habitat. The Habitat with high HSI value has low immigration rate (IR), and high emigration rate (ER) while habitat with low HSI value have high immigration rate, and low emigration rate (Simon, 2008). The HSI value of the habitat may increase due to immigration of species to low HSI, if not then species living in the habitats go extinct. The association between immigration rate (IR) and emigration rate (ER) is as shown in Fig. 2. For a linear association between rates and number of species, we have (Simon, 2008; MacArthur and Wilson, 1963, 1967)

$$\lambda_k = I\left(1 - \frac{k}{n}\right) \tag{1}$$

$$\mu_k = \left(\frac{E}{n}\right)k \tag{2}$$

Where $k = S$ and $n = S_{max}$ One can observe from Fig. 2, that the maximum immigration value is denoted λ and immigration rate by 'I'. Similarly, the maximum emigration value is denoted by μ and emigration rate by E. S and Smax represent the number of species and its maximum value. S_0 represents the number of species at equilibrium. We can observe that 'I' occurs when the island is empty ($k = 0$) and E occurs when no room is available for new species (at $n = S_{max}$) (MacArthur, 1972).

By assuming the special case, $I = E$, we have (Simon, 2008; MacArthur and Wilson, 1963; MacArthur and Wilson, 1967)

$$\lambda_k + \mu_k = E \tag{3}$$

Suppose at time t the island contains k species with probability P_k (t), the change in P_k (t) from t to ($t + \Delta t$), is explained as follows (Simon, 2008; MacArthur and Wilson, 1963; MacArthur and Wilson, 1967):

$$P_k(t + \Delta t) = P_k(t)(1 - \lambda_k\Delta t - \mu_k\Delta t) + P_{k-1}\lambda_{k-1}\Delta t + P_{k+1}\lambda_{k+1}\Delta t$$

The island contains 'k' species at time ($t+\Delta t$), if any one of the following conditions holds (Simon, 2008; MacArthur and Wilson, 1963; MacArthur and Wilson, 1967):

a. No immigration or migration during the interval Δt and k species at time t.
b. One species immigrated leaving (k-1) species at time t.
c. One species emigrated leaving (k + 1) species at time t.

Table 1
Extracted features.

Intensity based		Texture based		Shape based	
1	Average gray level	1	Energy	1	Area
2	Average contrast	2	Entropy2	2	Perimeter
3	Smoothness	3	Contrast	3	Compactness
4	Skewness	4	Mean	4	Normalized Standard deviation
5	Uniformity	5	Standard deviation	5	Area Ratio
6	Entropy1	6	Variance	6	Contour Roughness
		7	Correlation	7	Normalized Residual Value
		8	Homogeneity	8	Overlapping Ratio
		9	Sum average		
		10	Sum Variance		
		11	Sum entropy		

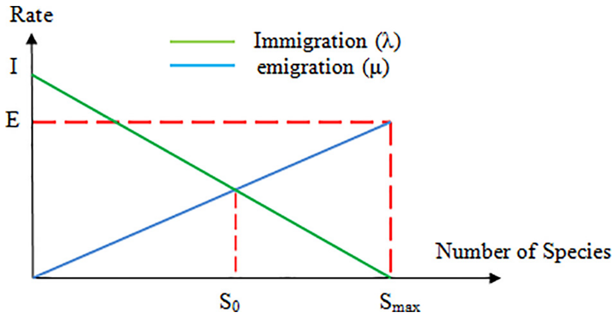

Fig. 2. Association between IR and ER.

Finally, we assume that Δt approaches to zero and produces-

$$P_k = \begin{cases} -(\lambda_k + \mu_k)P_k + \mu_{k+1}P_{k+1} + \lambda_{k-1}P_{k-1} 0 < k < n \\ -(\lambda_k + \mu_k)P_k + \mu_{k+1}P_{k+1} k = 0 \\ -(\lambda_k + \mu_k)P_k + \lambda_{k-1}P_{k-1} k = n \end{cases} \tag{4}$$

3.1.2. Biogeography-based optimization (BBO)

Biogeography-based optimization algorithm has some common features with population-based algorithms such as GA and PSO (Simon, 2008). One of the unique features of BBO is that it maintains elite solutions straight into the next generation. The proposed Biogeography-based optimization algorithm implemented for feature selection of masses in digital mammograms is as shown in Fig. 3. The BBO algorithm consists of the following steps-

3.1.2.1. Population encoding. In the proposed method, binary encoding technique is used for population encoding. In BBO, population is represented by habitat in the ecosystem. It consists of suitability index variables (SIV). Each feature belongs to one of the SIV. The habitat represents apossible solution of the feature selection problem. A variable that represents habitat is a 25-bit binary vector

```
Load feature data set        // 651 samples of 25 features
Initialize BBO parameters
    max          // maximum iterations
    psize        // number of habitats (population size)
    nf           // number of features
    mu           // Emigration Rate
    lamda        // Immigration Rate
    pm           // mutation coefficient
Define cost function Cost(f)  //  evaluate fitness
Generate initial population of habitas (i=1,2,3,……psize)
Encode the population using Binary encoding  technique
Compute  fitness of  initial population, Cost(f)= accuracy of ANFIS/ANN
Select Best_solutiion  from intial population
While (iteration<max)
  for i = 1 to psize
      for j= 1 to nf
         if rand< = lamda(i)  // Migration
           Compute emigration Probability EP
           Select habitat J using rouleteewheel selection based on EP
            Find migration solution
         end if
        if rand< = pm
         Apply Mutation on migration solution
        end if
     end loop j
     Gnerate  new solution
     Evaluate the fitness of new solution
   end  loop i
  Rank the solution and find current best
end of while
Obtain the final result as optimal features
```

Fig. 3. Biogeography-based optimization Algorithm.

representing twenty-five features extracted from segmented masses as shown in following example.

Feature Vector: 1101101011011010110110100

Value "1″ in the feature vector indicates that the corresponding feature is selected and a value "0" indicates that the feature is not selected.

3.1.2.2. Fitness calculation. The fitness of each habitat is represented by habitat suitability index (HSI). Classifiers ANFIS and ANN are used to ascertain fitness of each habitat in the population. Classification accuracy of these classifiers is used to find the fitness of each habitat. The fitness of ith habitat H is defined as –

$$fitness(H_i) = mean(Accuracy) \tag{5}$$

3.1.2.3. Migration. Migration improves poor islands by sharing the information of rich islands. The solutions with high fitness (HSI) value are good and they contain large species, while low HSI value represents few species. The migration solution is determined by the use of emigration rate (μ) and immigration rate (λ) probabilistically. Migration operation of biogeography-based optimization is the same as crossover operation of evolutionary algorithms. The only difference is that in evolutionary algorithms, new solution is generated by crossover operation while in BBO, the existing solution is updated by migration (Simon, 2008).

3.1.2.4. Mutation. Solution value (HSI) of the habitat may change due to natural catastrophic effect, diseases and flotsams. In BBO, this problem is modeled as mutation of the suitability index variable. The amount of mutation to be carried out is determined by probability of species count, as shown in Eq. (4). The habitat is selected for mutation using the roulette wheel selection method. Selection of the habitat is based on emigration probability.

The features selected by Biogeography-based optimization with adaptive neuro-fuzzy inference system (BBO-ANFIS) and Biogeography-based optimization with artificial neural network (BBO-ANN) are shown in Tables 2 and 3 respectively.

3.2. Classification

The proposed methodology uses Adaptive neuro-fuzzy inference system (ANFIS) and artificial Neural Network (ANN) for classification of masses.

3.2.1. Adaptive neuo- fuzzy inference system (ANFIS)

Adaptive neuro-fuzzy inference system is a combination of artificial neural network and fuzzy logic. Adaptive neuro-fuzzy inference system is a multilayer feedforward network, which uses supervised learning technique and fuzzy logic for input-output mapping (Jang and Sun, 1995). The primary mechanism used by fuzzy logic for mapping of input space to output space is the if-then rules. Fuzzy logic is a multi-value logic based on the concept of fuzzy sets. A fuzzy set contains elements with varying degree of membership between '0′ and '1′, represented by Ã. Fuzzy logic can change the qualitative parts of human information. Nevertheless, it does not have a characterized technique that can be utilized as a guide in the process of transformation of human idea into rule base fuzzy inference system (FIS). In addition, the time required to update membership functions (MFs) is quite high (Jang and Sun, 1995). Hence, ANN can be use to automatically modify MFs and minimize the rate of errors in the assurance of rules in fuzzy logic. Fuzzy inference system

The block diagram of fuzzy inference system is as shown in Fig. 4.

Fuzzy inference system consist of following elements-

Table 2
BBO-ANFIS based feature selection.

Max iteration	Pop size	Feature set	No. of Features selected	Selected Feature Numbers	Accuracy of feature selection (%)
10	10	FS1	15	2, 4, 5, 7, 8, 9, 10, 11, 14, 16, 17, 18, 19, 23, 24	97.70
15	10	FS2	14	2, 5, 6, 7, 8, 10, 11, 13, 15, 18, 22, 23, 24, 25	98.92
20	10	FS3	14	4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 22, 23, 24	98.00
25	10	FS4	13	1, 3, 4, 7, 8, 11, 17, 18, 19, 20, 21, 23, 25	96.16
30	10	FS5	13	2, 4, 6, 8, 9, 10, 14, 16, 17, 18, 22, 24, 25	96.16

Table 3
BBO-ANN based feature selection.

Max iteration	Pop size	Feature set	No. of Features selected	Selected Feature Numbers	Accuracy of feature selection (%)
10	10	NN1	11	1, 5, 6, 7, 8, 14, 20, 21, 23, 24, 25	95.85
15	10	NN2	12	4, 5, 6, 8, 9, 11, 12, 16, 17, 18, 20, 23	95.08
20	10	NN3	13	2, 4, 6, 8, 10, 15, 16, 18, 20, 22, 23, 24, 25	95.54
25	10	NN4	13	1, 4, 6, 8, 10, 11, 13, 14, 15, 16, 20, 21, 25	94.16
30	10	NN5	11	2, 5, 6, 8, 15, 16, 17, 19, 20, 23, 25	95.69

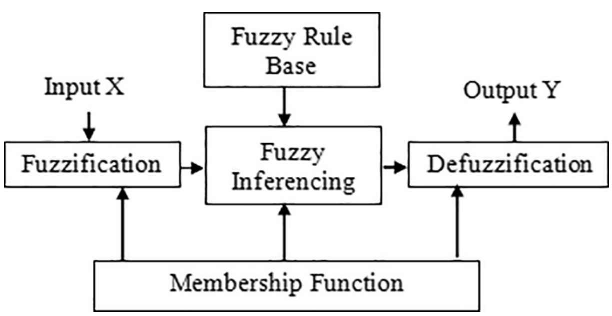

Fig. 4. Fuzzy inference system.

Input Vector: X = [x_1, x_2,... x_n] ^{T} are crisp values, which are transformed into fuzzy sets using Fuzzification process.

Output Vector: Y = [y_1, y_2,... y_m] ^{T} are the crisp values which are transformed using defuzzification process i.e. fuzzy values are transformed into crisp values.

Membership function: A membership function specifies the degree to which a given input belongs to a set. The membership function associated with a fuzzy set is denoted by $\mu\tilde{A}(x)$. The value $\mu\tilde{A}(x)$ is called membership degree of x in set Ã and it lies between 0 and 1. It is used in the Fuzzification and defuzzification steps of a FLS (fuzzy logic system), to map the non-fuzzy input values to fuzzy linguistic terms and vice versa. There are different forms of membership functions such as, Triangular, Trapezoidal, Piecewise linear, Gaussian and Singleton.

Fuzzification: It is a process of transforming crisp values into grades of membership (fuzzy values between 0 and 1) for linguistic terms, "far", "near", "small" of fuzzy sets.

Fuzzy Rule base: The primary mechanism used by fuzzy logic for mapping of input space to output space is if-then rules. A fuzzy rule base is a collection of propositions containing linguistic variables. The rules are expressed in the form-

If x is A and y is B then z is C where x, y and z represent variables and A, B and Z are linguistic variables (e.g. 'far', 'near', 'small').

Fuzzy Inferencing: It combines facts obtained from the Fuzzification with the rule base and conducts the Fuzzy reasoning process.

Defuzzification: It is the process of translating fuzzy set values back to the real world values (crisp).

There are several types of FIS, namely Takagi–Sugeno, Mamdani, and Tsukamoto. FIS of Takagi–Sugeno (Takagi and Sugeno, 1985) model is widely used in the application of ANFIS method.

3.2.1.1. ANFIS architecture. Adaptive network is one example of feedforward neural network based on supervised learning (Jang, 1993). ANFIS is functionally similar to first-order Takagi-Sugeno fuzzy inference system (Takagi and Sugeno, 1985). For simplicity, we assume that the fuzzy model consists of two inputs and one output, defined as *a*, *b* and *c* respectively. For first-order Takagi–Sugeno fuzzy model the rule base consists of two fuzzy if-then rules (Jang and Sun, 1995).

Rule 1: if a is X_1 and b is Y_1 then c is $r_1a+s_1b+t_1$

Rule 2: if a is X_2 and b is Y_2 then c is $r_2a+s_2b+t_2$where r_i, s_i and t_i (i = 1 or 2) are linear parameters.

The architecture of ANFIS is as shown in Fig. 5, and consists of five layers.

The nodes at 1st and 4th layers are represented by square symbols called adaptive nodes and they consist of parameters. Similarly, nodes at 2nd and 3rd layer are represented by circles called fixed node and they do not consist of any parameter. The parameters of adaptive nodes are updated by a training algorithm to achieve the desired output. The nodes at 1st layer are responsible for generating fuzzy membership values for input data. Usually bell-shaped fuzzy membership function specifies the fuzzy set. Parameters used in the first layer are called, premise parameters. The nodes at 2ndlayer uses fuzzy AND operator. The output of each node at 2nd layer represents the firing strength of the rule. The nodes at 3rd layer scale the firing strength. Output generated by this layer is normalized firing strength. The nodes at 4th layer generate output using normalized firing strength from previous layer nodes and ($r_ia + s_ib + t_i$) parameter at this node. The parameters used at this layer are called consequent parameters. The 5th layer consists of a single node. It is a summation unit. It generates an output by summing up all the inputs received from the previous layer.

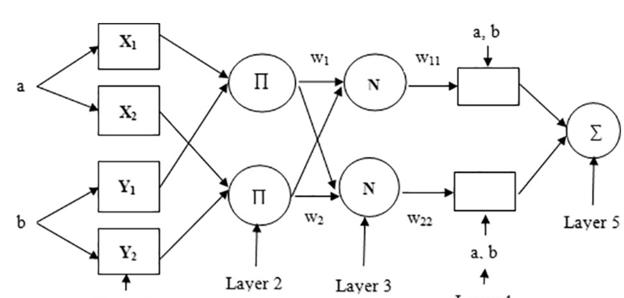

Fig. 5. Architecture of Adaptive neuro-fuzzy inference system.

Please cite this article in press as: Thawkar, S., Ingolikar, R. Classification of masses in digital mammograms using Biogeography-based optimization technique. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.01.004

3.2.2. Artificial neural network(ANN)

An Artificial Neural Network is a computational model based on the structure and function of the biological neuron system. Artificial neural network is a parallel-distributed system consisting of highly interconnected neural computing elements called neurons (artificial), which have the ability to learn and acquire knowledge. This is made use of to solve problems (Bovis et al., 2000; Baeg and Kehtarnavaz, 2000; Velthuizen and Gaviria, 1999). ANN is trained by supervised or unsupervised training algorithm. ANN has been widely used in the fields of pattern recognition, optimization, image processing and forecasting, which are tough to solve using normal rule-based methods.

The proposed classification method uses multilayer perceptron (MLP) for the classification of masses. It consists of three layers: input layer, hidden layer and output layer (Cheng et al., 2006). In multilayer perceptron, information flows in only the forward direction; that is why it also is called feedforward neural network. The multilayer perceptron net was trained using supervised learning method. Training of net requires input-output samples. The output generated by the network in response to training data is, compared with the target data for calculation of error. An error determines the amount of weights change for the network. The network is trained repeatedly with input data and weights updated until input-output mapping occurs. In ANN, mean square error (MSE) determines the amount of weight change. Once the network trained, it is tested with unseen data, to evaluate the performance of the model.

4. Results and discussion

The proposed methodology evaluated over 651 digital mammograms obtained from Digital Database for Screening Mammography (DDSM). It is available at source www.marathon.csee.usf.edu/mammography/Database.htm (Heath et al., 1998; Heath et al., 2001). Out of 651 cases, 314 were benign and 337 were malignant. As discussed in Section 3.1.2, Biogeography-based optimization with ANFIS and ANN selected optimal features from the set of twenty-five features. Parameter values used in Biogeography-based optimization were-

Number of Variables (features):	25
Max. Number of Generations:	10/15/20/25/30
Population size of habitats:	10
Mutation rate:	0.05

These parameters were selected by trial and error method, except the number of variables and maximum number of habitats. The parameter value of the maximum number of generations initially was tested for 10, 25, 50 and 100 respectively, and the best results were found for values below 50. Similarly, mutation rate was tested for 0.1 and 0.05, and the best results were found for value 0.05.

Selection of optimal features depends on the fitness value of the habitat. The classification accuracy of ANFIS and ANN acts as fitness value of the habitat as defined in Eq. (5). The features selected by BBO-ANFIS are shown in Table 2. From Table 2, one can observe that five feature subsets were selected by BBO-ANFIS. The subset FS2 appears to be best with an accuracy of 98.92% having 14 features. Similarly, features selected by BBO-ANN are shown in Table 3. The feature subset NN1 appears to be best with an accuracy of 95.85% having 11 features. The relation between emigration rate (ER), immigration rate (IR) and emigration probabilities (EP) for feature selection using BBO-ANFIS and BBO-ANN with population size of 10, is shown in Fig. 6. From Fig. 6, we observe that emigration rate decreases from 1 to 0, while immigration rate

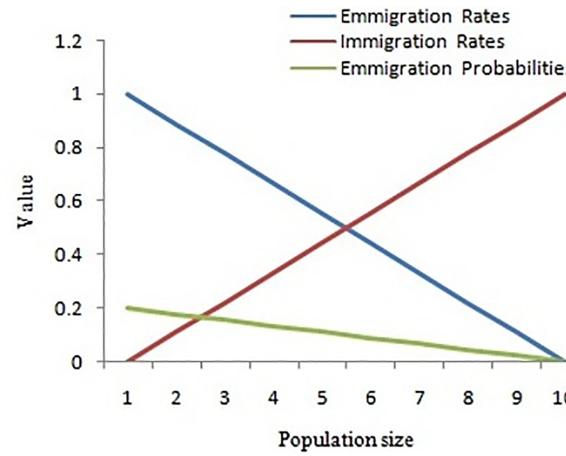

Fig. 6. Relation between IR, ER and EP.

increases from 0 to 1, with population size. Similarly, emigration probability decrease from 0.2 to 0 with population size.

The feature subset FS2 selected by BBO-ANFIS, is used to train and test adaptive neuro-fuzzy inference system classifier. The training data set FS2 consist of 651 samples of 14 features, as shown in Fig. 7.

The parameters used for training and testing of multi-model ANFIS classifier are as shown below:

Name:	'anfis'
Type:	'sugeno'
Train Data:	651 samples
Train Class:	benign/malignant
And Method:	'prod'
Or Method:	'max'
Defuzzification Method:	'wtaver'
Implication Method:	'prod'
Aggregation Method:	'max'
Number of Epoch:	100
Number of membership functions:	3
Input Membership function:	'gbellmf'
Output Membership function:	'linear'

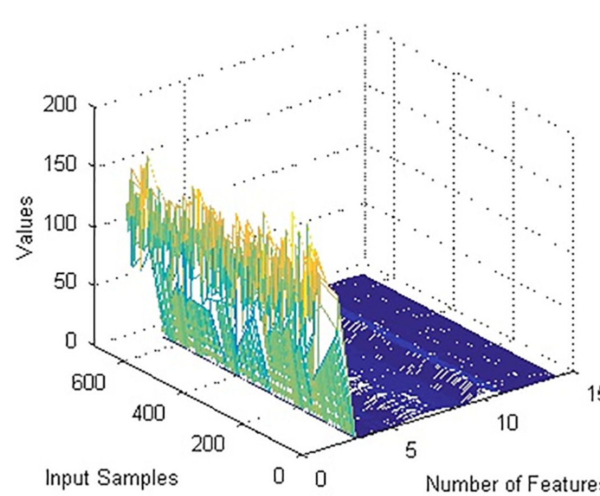

Fig. 7. ANFIS training Data.

Please cite this article in press as: Thawkar, S., Ingolikar, R. Classification of masses in digital mammograms using Biogeography-based optimization technique. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.01.004

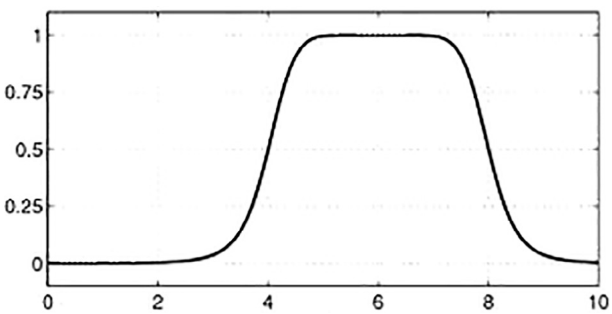

Fig. 8. Membership function gbellmf.

The generalized bell ('gbellmf') membership function as shown in Fig. 8, depends on three parameters *a*, *b*, and *c* as given by

$$f(x;a,b,c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \tag{6}$$

The bell ('gbellmf') membership function consists of one extra parameter than the Gaussian membership function. The membership function will approach a non-fuzzy set if the extra parameter is tuned. The 'linear' membership function is associated with the output. The proposed ANFIS is a Sugeno-type system consisting of only one output. The membership function associated with output is either linear or constant function.

ANN (Multilayer Perceptron) was trained and tested using feature subset NN1. The training data set NN1 consists of 651 samples of 11 features as shown in Fig. 9.

The input data consists of 70% part of training and the remaining 30% is part of testing. The proposed multilayer perceptron model consists of 11 nodes at input layer, 12 nodes at hidden layer and 2 nodes at output layer. The network was trained using Bayesian regularization method that updates the weight and bias values according to Levenberg-Marquardt optimization. ANN trained for 1000 epochs. The weights and bias of the network has updated according to mean square error (MSE). The values of network parameters obtained at 1000 epoch are, network performance 0.1474, gradient 0.001552, Mu (momentum) 50, number of parameters 40.023 and sum square performance 435.77.

Statistical parameters as listed in Table 4 were used to evaluate the performance of classifiers. The performance summary of ANFIS and ANN classifiers is as shown in Table 5. From Table 5, one can observe that ANFIS is superior to ANN, when we evaluate the performance of classifiers with all the statistical parameters listed in Table 4. ANFIS classifier achieved the highest classification accuracy of 98.92% with 99.1% sensitivity and 98.72% specificity, while artificial neural network achieved an accuracy of 95.85% with sensitivity and specificity of 97.92% and 93.63% respectively. The misclassification rate of adaptive neuro-fuzzy inference system and artificial neural network classifier was 1.08% and 4.15% respectively. Similarly, false alarm or type-I error (FPR) and type-II error (FNR) for ANFIS was 1.27% and 0.89% respectively, while for ANN it was 6.36% and 2.077% respectively.

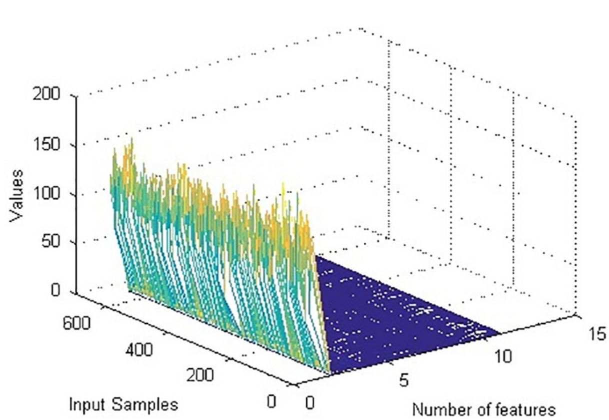

Fig. 9. ANN training Data.

Table 4
Parameters for measuring performance of classifiers.

$Accuracy(ACC) = (TP+TN)/(TP+TN+FP+FN)$
$Sensitivity(TPR) = TP/(TP+FN)$
$Specificity(TNR) = TN/(TN+FP)$
$Type-I\ error\ (FPR) = FP/(FP+TN)$
$\mathrm{Type-II\ error\ (FNR) = FN/(FN+TP)}$
$Mean\ square\ error\ (MSE) = \frac{1}{n}\sum_{1}^{n}(O_i - T_i)^2$
$Root\ mean\ square\ error\ (RMSE) = \sqrt{MSE}$
TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

Table 5
Classification Results of ANFIS and ANN.

Method	TP	FN	TN	FP	TPR (%)	TNR%)	ACC (%)
ANFIS	334	3	310	4	99.10	98.72	98.92
ANN	330	7	294	20	97.92	93.63	95.85

Correlation coefficient (R), mean square error (MSE) and root mean square error (RMSE) measure the predictive power of the model. The relationship between output and target is linear if R-value is '1', and the model is 100% correct. The R-value of ANFIS for training and testing is 0.8826 and for ANN it is 0.8648. As far as MSE is concerned, the smaller the value better is the model. MSE of ANFIS and ANN classifier per input sample areas shown in Figs. 10 and 11 respectively. From Figs. 10 and 11, we can observe that MSE of ANFIS and ANN per input sample varies between -0.184 to 1.233, and 0.0674 to 0.625, respectively. The mean MSE value for ANFIS and ANN is 0.0607 and 0.1477 respectively. Similarly, the root mean square error (RMSE) is defined as the square root of differences between target and output generated by network. The Mean RMSE value for ANFIS and ANN is 0.2465 and 0.3757 respectively. Statistical results show that ANFIS classifier is better than ANN classifier.

One more useful parameter used to measure the performance of the classifiers is the area under receivers operating (ROC) curve. The ROC curve is a graph between true positive rate (TPR) and false positive (FPR) rate i.e., 1-Specificity. Its value lies between '0' and '1'. The model or classifier is said to be 100% correct if its value is equal to '1' (Swets, 1988). The ROC curve and area under curve for ANFIS and ANN are as shown in Fig. 12 and Table 6 respectively.

From Table 6, we can observe that area under ROC curve with 95% Confidence Interval (C.I.) for ANFIS is $A_Z = 0.999 \pm 0.000$ and for ANN is $A_Z = 0.966 \pm 0.008$. From Fig. 12, it is clear that ROC curve for both ANFIS and ANN is close to one.

The outcomes of the proposed BBO-ANFIS method was compared with existing studies as shown in Table 7. The comparative study shows that the proposed BBO-ANFIS method is far better than the existing one. The performance of the classifier was greatly affected by the use of few and unbalanced data (Kubat and Matwin, 1997; Kubat et al., 1998). The proposed work uses large and balanced data set of benign and malignant cases.

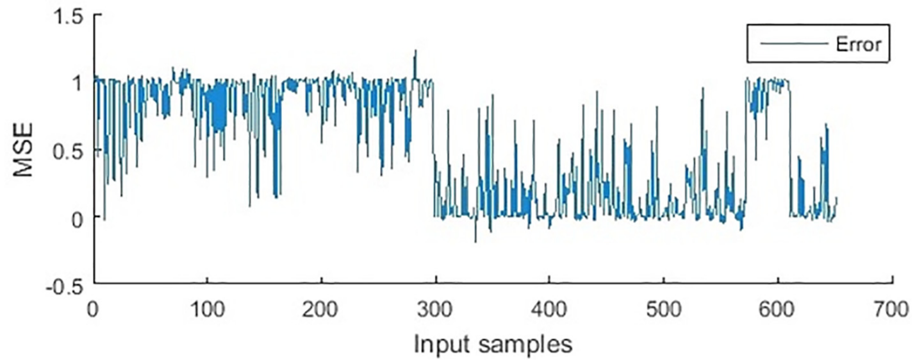

Fig. 10. Mean square error of ANFIS classifier.

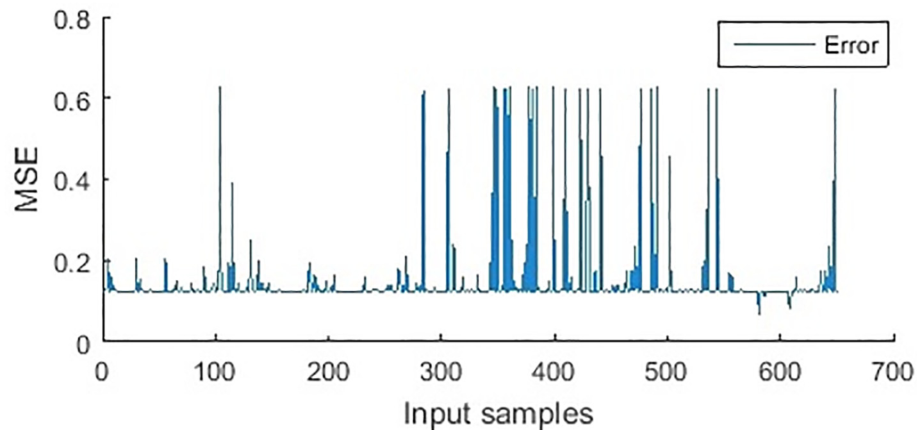

Fig. 11. Mean square error of ANN classifier.

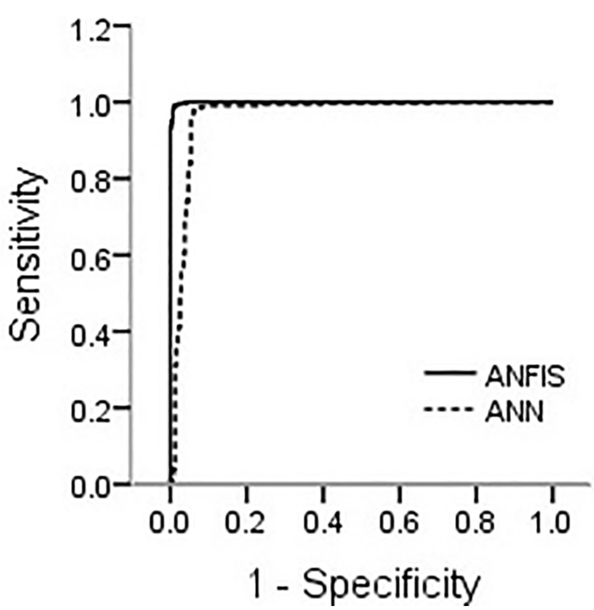

Fig. 12. Comparison of ROC curves of ANFIS and ANN.

Table 6
Area under curve for ANFIS and ANN.

Classifiers	Area	Standard Error	Asymptotic Sig.	95% C. I.	
				LB	UB
ANFIS	0.999	0.000	0.000	0.999	1.000
ANN	0.966	0.008	0.000	0.950	0.983

Table 7
Comparison of proposed method with existing work.

Authors	Algorithm	Classification Problem	Accuracy (%)
Zhang et al., 2016a	BBO- FNN	Fruit	89.11
Zhang et al., 2016b	MLP-BBO	Masses	92.52
Khehra and Pharwaha, 2017	BBO-SVM	Microcalcifications	93.36
Huang et al., 2012	PSO-ANFIS	Masses	92.8
Proposed	BBO-ANFIS	Masses	98.92

5. Computational complexity

Finally, we consider the computational complexity of proposed Biogeography based optimization algorithm discussed in Fig. 3. The complexity of an algorithm plays a major role in design of efficient algorithms.

Let us assume N = maximum number of iterations, M = population size and NF = number of features then the complexity of line No. 8, 9, and 10 is O(N∗M∗NF). If we consider N = M = NF then complexity is $O(N^3)$. The complexity of Roulette wheel selection (line No. 13) is O(N). Then total time required to select a habitat J using roulette wheel selection is $O(N*N)$. Let f be the complexity of fitness function. Then the computational complexity of algorithm is $O(N^3 + (N*N) + N^2*O(f))$. The proposed method was implemented in MATLAB R2015a and executed on Pentium(R) Dual-Core E5700@3 GHz processor with 1 GB RAM. An algorithm takes average execution time of 18 s/image for automatic detection of masses in digital mammograms.

With the above discussions, we conclude that BBO-ANFIS appears to be the best methodology. It can help to improve breast cancer diagnosis. However, this accuracy achieved with BBO-ANFIS is at the expense of computational complexity.

6. Conclusion

This paper proposes an efficient technique for feature selection and classification of masses in digital mammograms. Biogeography-based optimization technique with ANFIS and ANN is used for feature selection. The features selected by BBO-ANFIS and BBO-ANN are used to train and test ANFIS and ANN classifiers. Performance analysis shows that BBO-ANFIS is superior to BBO-ANN. Although BBO-ANFIS is superior to BBO-ANN, its computation time is higher as compared to BBO-ANN. Results achieved with the proposed method are far better as compared to existing methods. The suggested method can help to improve breast cancer diagnosis and would minimize mortality rate. In future work, we would concentrate on improving computation time of the feature selection and classification system.

7. Declarations of interest

'none'

References

Acharya, U.R., Fujita, H., Sudarshan, V.K., Oh, S.L., Adam, M., Tan, J.H., Chua, K.C., 2017. Automated characterization of coronary artery disease, myocardial infarction, and congestive heart failure using contourlet and shearlet transforms of electrocardiogram signal. Knowl.-Based Syst. 132, 156–166.
Baeg, S., Kehtarnavaz, N., 2000. Texture based classification of mass abnormalities in mammograms. Computer-based medical systems, 2000. CBMS 2000. In: Proceedings. 13th IEEE Symposium on, pp. 163–168.
Bovis, K., Singh, S., Fieldsend, J., Pinder, C., 2000. Identification of masses in digital mammograms with MLP and RBF\nnets. In: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, 1, pp. 342–347.
Chang, F., Chang, Y., 2006. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Adv. Water Resour. 29 (1), 1–10.
Cheng, H., Shi, X., Min, R., Hu, L., Cai, X., Du, H., 2006. Approaches for automated detection and classification of masses in mammograms. Pattern Recogn. 39 (4), 646–668.
DeSantis, C., Siegel, R., Bandi, P., Jemal, A., 2011. Breast cancer statistics, 2011. CA: Cancer J. Clinicians 61 (6), 408–418.
Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr, P., Moore, R., Chang, K., Munishkumaran, S., 1998. Current status of the digital database for screening mammography. Digital Mammography, 457–460.
Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, P., 2001. The digital database for screening mammography. Proc. Fifth Int. Workshop Digital Mammography, 212–218.
Huang, M.-L., Hung, Y.-H., Lee, W.-M., Li, R., Wang, T.-H., 2012. Usage of case-based reasoning, neural network and adaptive neuro-fuzzy inference system classification techniques in breast cancer dataset classification diagnosis. J. Med. Syst. 36 (2), 407–414.
Jang, J., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. 23 (3), 665–685.
Jang, J., Sun, C., 1995. Neuro-fuzzy modeling and control. Proc. IEEE 83 (3), 378–406.
Khehra, B., Pharwaha, A., 2017. Comparison of Genetic Algorithm, Particle Swarm Optimization and Biogeography-based Optimization for Feature Selection to Classify Clusters of Microcalcifications. J. Inst. Eng. (India): Ser. B 98 (2), 189–202.
Kubat, M., Holte, R.C., Matwin, S., 1998. Machine learning for the detection of oil spills in satellite radar images. Mach. Learn. 30 (2–3), 195–215.
Kubat, M., Matwin, S., 1997. Addressing the curse of imbalanced training sets: one sided selection. Icml 97, 179–186.
Lahmiri, S., Boukadoum, M., 2013. Hybrid discrete wavelet transform and gabor filter banks processing for features extraction from biomedical images. J. Med. Eng., 1–13
Lahmiri, S., Boukadoum, M., 2011a. Hybrid discrete wavelet transform and Gabor filter banks processing for mammogram features extraction. In: 2011 IEEE 9th International New Circuits and Systems Conference, NEWCAS 2011. pp. 53–56.
Lahmiri, S., Boukadoum, M., 2011b. DWT and RT-based approach for feature extraction and classification of mammograms with SVM. In: 2011 IEEE Biomedical Circuits and Systems Conference, BioCAS 2011. pp. 412–415.
Li, H., Wang, Y., Liu, K., Lo, S., Freedman, M., 2001. Computerized radiographic mass detection - part ii: decision support by featured database visualization and modular neural networks. IEEE Trans. Med. Imaging 20 (4), 302–313.
MacArthur, R.H., 1972. Geographical ecology: patterns in the distribution of species. Princeton University Press.
MacArthur, R., Wilson, E., 1963. An equilibrium theory of insular zoogeography. Evolution 17 (4), 373.
MacArthur, R.H., Wilson, E.O. Theory of Island Biogeography. Princeton, NJ.
Mousa, R., Munib, Q., Moussa, A., 2005. Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. Expert Syst. Appl. 28 (4), 713–723.
Raghavendra, U., Rajendra Acharya, U., Fujita, H., Gudigar, A., Tan, J.H., 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. J. 46, 151–161.
Raghavendra, U., Fujita, H., Gudigar, A., Shetty, R., Nayak, K., Pai, U., Acharya, U.R., 2018. Automated technique for coronary artery disease characterization and classification using DD-DTDWT in ultrasound images. Biomed. Signal Process. Control 40, 324–334.
Sameti, M., Ward, R., Morgan-Parkes, J., Palcic, B., 1997. A method for detection of malignant masses in digitized mammograms using a fuzzy segmentation algorithm. In: Proceedings of the 19th Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vol 19, Pts 1–6, 19, 513–516.
Simon, D., 2008. Biogeography-based optimization. IEEE Trans. Evol. Comput. 12 (6), 702–713.
Swets, J., 1988. Measuring the accuracy of diagnostic systems. Science (New York, N. Y.) 240 (4857), 1285–1293.
Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its applications to modeling and control. Syst. Man Cybern. IEEE Trans. SMC-15 (1), 116–132.
Thawkar, S., Ingolikar, R., 2017a. Automatic detection and classification of masses in digital mammograms. Int. J. Intell. Eng. Syst. 10 (1), 65–74.
Thawkar, S., Ingolikar, R., 2017b. Efficient approach for the classification of masses in digital mammograms. Int. J. Innovative Comput. Inf. Control 13 (3), 967–978.
Velthuizen, R.P., Gaviria, J.I., 1999. Computerized mammographic lesion description. In: Proceedings of the First Joint BMES/EMBS Conference, 1999, 2, 1034-vol.
Wesche, T., Goertler, C., Hubert, W., 1987. Modified Habitat Suitability Index Model for Brown Trout in Southeastern Wyoming. North Am. J. Fish. Manag. 7 (2), 232–237.
Zhang, Y., Phillips, P., Wang, S., Ji, G., Yang, J., Wu, J., 2016a. Fruit classification by biogeography-based optimization and feedforward neural network. Expert Syst. 33 (3), 239–253.
Zhang, Y., Wu, X., Lu, S., Wang, H., Phillips, P., Wang, S., 2016b. Smart detection on abnormal breasts in digital mammography based on contrast-limited adaptive histogram equalization and chaotic adaptive real-coded biogeography-based optimization. Simulation 92 (9), 873–885.

Please cite this article in press as: Thawkar, S., Ingolikar, R. Classification of masses in digital mammograms using Biogeography-based optimization technique. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.01.004