A MODIFIED BINARY PSO BASED FEATURE SELECTION FOR AUTOMATIC LESION DETECTION IN MAMMOGRAMS

Sheba K.U[1], Gladston Raj S[2] and Ramachandran D[3]

[1]Department of Computer Applications, BPC College, Piravom
[2]Department of Computer Science, Government College, Nedumangad
[3]Professor and HOD, Department of Imageology, Regional Cancer Center, Thiruvananthapuram

ABSTRACT

This paper presents an effective feature selection method that can be applied to build a computer aided diagnosis system for breast cancer in order to discriminate between healthy, benign and malignant parenchyma. Determining the optimal feature set from a large set of original features is an important pre-processing step which removes irrelevant and redundant features and thus improves computational efficiency, classification accuracy and also simplifies the classifier structure. A modified binary particle swarm optimized feature selection method (MBPSO)has been proposed where k-Nearest Neighbour algorithm with leave-one-out cross validation serves as the fitness function. Digital mammograms obtained from Regional Cancer Centre, Thiruvananthapuram and the mammograms from web accessible mini-MIAS database has been used as the dataset for this experiment. Region of interests from the mammograms are automatically detected and segmented. A total of 117 shape, texture and histogram features are extracted from the ROIs. Significant features are selected using the proposed feature selection method.Classification is performed using feed forward artificial neural networks with back propagation learning. Receiver operating characteristics (ROC) and confusion matrix are used to evaluate the performance. Experimental results show that the modified binary PSO feature selection method not only obtains better classification accuracy but also simplifies the classification process as compared to full set of features. The performance of the modified BPSO is found to be at par with other widely used feature selection techniques.

KEYWORDS

Binary particle swarm optimization, Feed forward artificial neural networks, Feature selection, k-Nearest Neighbour.

1. INTRODUCTION

Breast Cancer is the most common cancer affecting women across the world [1].In India there is a rising incidence of breast cancer especially among young women. Around 48% of breast cancer patients in India are below the age of 50 [2]. 60% of the breast cancer cases in India are diagnosed at an advanced stage due to the lack of breast cancer awareness, lack of screening facilities and due to incorrect diagnosis. This drastically affects survival rate and treatment options [2].

Currently, the most widely accepted screening modality for breast cancer is mammography as it is reliable and economical [3]. Space occupying lesions are the most common symptoms of breast cancer in mammograms [4]. Space occupying lesions can be of three types- masses, asymmetrical

breast tissue and architectural distortion of the breasts [5]. All these lesions can be classified as either benign or malignant depending on their shape, texture, density and grey level intensity values. Hence, each mammogram requires detailed evaluation in order to differentiate healthy, benign and malignant parenchyma. Chances are that malignancies in mammograms may go undetected or can be diagnosed incorrectly due to the strenuous job of evaluation, poor quality of mammograms and subtle nature of malignancies [6]. A computer aided detection and diagnosis system (CAD) for breast cancer can aid the radiologist in interpreting the mammograms and help in the detection of suspicious lesions. They can provide a second opinion while the final decision lies with the radiologist. Recent studies have shown that computer aided detection and diagnosis systems have helped greatly in improving the radiologists' accuracy in interpreting mammograms [7].

The efficiency of a CAD system greatly depends on its accuracy, computational time and ease of use. At present, the accuracy of CAD systems is not very high [8]. But, there is a need for near perfection in lesion detection and diagnosis by the CAD systems. This is because false positive rate can create undue anxiety and stress among patients whereas false negative rates can prevent early detection of breast cancer which can cause serious threat to the patient's life. Hence, improving the accuracy of CAD system is very important as it can aid in improving the diagnostic decisions.

CAD systems involve the following phases- Image acquisition, Image pre-processing, Image segmentation, Feature extraction, Feature selection and Classification [9]. The performance of CAD systems depend more on the optimization of feature subset selection than on the classification methods. In mammograms, there is a large variation in the appearance of normal, benign and malignant breast tissues with respect to their shape, texture and grey level intensity values. Hence, it is necessary to extract texture, shape and grey level intensity features from the ROIs. [10]. As a result, a large, diverse and complex feature set is obtained. Not all features are required for classification as some of them are redundant, irrelevant, noisy and misleading and can actually degrade the performance of the classifier. Also, if the training data set is small when compared to the size of feature set, it can lead to the situation called curse of dimensionality [11]. This can also reduce the classifier performance. Due to these reasons, feature selection is necessary to obtain optimal subset of features that can maximize classification accuracy and reduce running time.

The aim of the paper is to develop an effective feature selection method that guarantees the selection of optimal features which can greatly improve the performance of the classifier and reduce its computational time.

2. OVERVIEW OF FEATURE SELECTION METHODS

Feature selection methods have been categorized into two types- filter and wrapper methods [12].Filter method chooses an optimal subset of features by eliminating less significant features using the statistical properties of the features. Wrapper approach incorporates learning algorithm to select optimal subset of features. The wrapper approach usually outperforms the filter approach in terms of classification accuracy as the former selects the optimal subset of features based onthe performance of the feature subset when applied on a classification algorithm [13].

Due to limitations in conventional feature selection methods, recent research have employed evolutionary computational techniques for feature selection. They include particle swarm optimization (PSO), genetic algorithm, ant colony optimization etc. These techniqueshave been extensively used in feature selection. PSO [14] has been used by many researchers for feature selection as it has global searching ability, is easy to implement, converges quickly and takes less

computation time[15]. A number of recent studies have focused on PSO based feature selection methods.

A. Unler et al. [13] in their paper have proposed a modified discrete PSO algorithm for feature selection and compared it with tabu search and scatter search algorithms using publicly available datasets. The algorithm was found to be competitive in terms of classification accuracy and computational performance.

Xue et al.[16] in their paper developed a feature selection method based on modified binaryparticle swarm optimization method. Decision tree classifier has been used for classification.It has been compared with two traditional features selection methods by applying it on 14 benchmarkproblems of varying difficulty. This method is found to achieve better performance compared to the traditional feature selection methods.

A review of PSO algorithms and their variations are presented by the authors Tran et al. in their paper [17]. Current issues and challenges for future research are also discussed.

In their paper [18], authors Yong et al. developed the barebones PSO to find optimal feature subset where a reinforced strategy is designed to update the local leaders of particles in order to avoid degradation of outstanding genes in particles. 1-NN is used as the classifier to evaluate the performance and experiments show that the algorithm is competitive in terms of accuracy and computational performance.

The authors Wong et al. [19] proposed an effective technique to classify regions of interest(ROIs) of digitized mammograms into mass and normal breast tissue region by using PSO based feature selection and SVM classifier. This method was successful in finding significant features that greatly improved the classification accuracy of SVM.

In paper [20], the authors have proposed a modified PSO based feature selection for classification of lung CT images. The experimental results shows higher classification accuracy compared to basic PSO feature selection method.

The authors Zyoutet al. in their paper [21] have used PSO-kNN to select relevant GLCM features for classification of microcalcification clusters in digital mammograms. They have obtained a class accuracy of 88% which reveals that feature selection using PSO-kNN is effective.

Though extensive research has been done using PSO based feature selection in the classification of microcalcifications in mammograms, it is found that not much research using PSO based feature selection has been done in digital mammograms for classification of lesions as benign or malignant [22].

In this paper, we propose a modified binary particle swarm optimization (MBPSO) algorithm for selection of optimal feature subset for classification of mammograms as healthy, benign or malignant. The modified binary particle swarm optimization introduced in this paper belongs to the wrapper approach category. The modified version of BPSO differs from the traditional BPSO in two aspects. They are:

1. In the traditional BPSO, velocity update is calculated from the previous velocity using two best values- local best(lbest) and global best (gbest).In the modified version of BPSO presented in this paper, besides lbest and gbest, iteration best(Itbest) is also considered for velocity update .It is the best position obtained among particles in each iteration.

2. Secondly, in case the best fitness value is shared by two or more particles, the positions of particle containing less number of features will be taken as the best position for lbest, gbest and Itbest.

k-Nearest Neighbor (KNN) with leave-one-out method is used as the fitness function to choose the optimal feature subset in MBPSO. Feed forward artificial neural networks (FFANN) with back propagation has been used as the classifier. FFANN is trained using the optimal feature subset. After necessary accuracy is obtained, the weights are frozen. Test data is then fed to the FFANN and classification accuracy is measured.

The rest of the paper is organized as follows. Section 3 presents the traditional PSO and BPSO method. The proposed methodology is described in section 4. Section 5 provides experimental results and performance analysis. Section 6 concludes the paper.

3. BINARY PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization (PSO) is a population based search technique for finding optimal solution in real number space modeled after the social behavior of bird flocks [23]. The concept developed by Kennedy and Ebenhart consists of the following steps.

1. Initialize a set of random potential solutions called particles each of which are assigned random position X_i and velocity V_i on D dimensions.

2. Evaluate the fitness function of each particle i in D dimensions. If the current fitness value is better than the earlier fitness value obtained by the particle i, the local best value $lbest_i$ of the particle i is updated to the current fitness value. The current location $X_i = \{x_{i1}, x_{i2...} x_{iD}\}$ is assigned as the local best position $lb_i = \{lb_{i1}, lb_{i2} \ldots\ldots lb_{iD}\}$.

3. Identify the particle with best fitness value achieved so far in the entire swarm. The position of that particle is assigned as the global best position and is represented as $G = \{g_1, g_{2....} g_D\}$. The best fitness value of that particle is assigned as the global best value gbest.

4. The position X_i and the velocity V_i of each particle is updated using the following equation.
   $V_i = \omega V_i + c_{1*} r_{1*} (lb_i - X_i) + c_{2*} r_{2*} (G - X_i)$ (1)
   $X_i = X_i + V_i$ (2)
   where c_1 and c_2 are learning rates; r_1 and r_2 are random numbers in the range [0,1]; ω is the inertia weight.

5. Loop steps 2-4 until a good fitness value G is attained or a maximum number of iterations are reached.

The original PSO was designed for real valued problems operating in continuous space. Since many problems occur in discrete space, the original authors extended the real- valued version of PSO to binary/ discrete space and named it as binary particle swarm optimization(BPSO) [24]. Since feature selection is based on discrete qualitative differentiation between variables, BPSO is found to be more apt for feature selection [17]

There are two main differences between original PSO and BPSO. They are,

1. Particles in BPSO are represented as binary vectors i.e. as 0's and 1's. If x_{ij}=1, the feature j of particle i has been selected, if x_{ij}=0, the feature is not selected.

2. In BPSO, the velocity is treated as probability vector which determines whether a binary variable should take the value 0 or 1. Velocity is calculated in the same manner as in PSO. It is converted to probability vector in the range (0, 1) using a sigmoid function. It is given as

$$S_{ij}=\frac{1}{1+e^{-vij}} \tag{3}$$

The position of the j^{th} bit in the i^{th} particle is updated using S_{ij}as follows.

$$x_{ij} = \begin{cases} 1 & if\ \ \delta < S_{ij} \\ 0 & if\ otherwise \end{cases} \tag{4}$$

where δ is a random number between 0 and 1.

4. MODIFIED BINARY PARTICLE SWARM OPTIMIZATION METHOD

During the initial implementation of traditional BPSO to obtain the optimal feature subset,it was observed that the performance of the BPSO can be further improved if the following changes are incorporated in the algorithm.

- It was observed that when a particle bit, its local best bit value and global best bit value are all same, it can lead to a state where the probability to include or exclude the feature is 0.5. In problems with large number of features, it can result in high diversification. To overcome this problem,a modification was made to the equation for velocity updation by including a new factor called Iteration best(Itbest). Itbest is the best position attained by any particle in each iteration.

$$V_{id}=\omega * v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (It_d - x_{id}) + c_3 r_3 (g_d - x_{id}) \tag{5}$$

Here c_3 is the learning rate for the best position in each iteration. r_3 is a random number uniformly distributed in [0,1].

- During the execution of the algorithm, it is found that the fitness value generated for a particle in a particular iteration may be equal to its local best value or the global best value or the Iteration best value. In such cases, numbers of features are also considered in choosing the best solution. If the current fitness value calculated for particle i happens to be equal to it its local best value, but the number of features used to calculate the current fitness value is less than the number of features used to calculate the local best value, then the current position of particle i is chosen as the local best position and the current fitness value is updated as the local beat value.

i.e. $|x_i| < |lbest_i|$

Then $(lb_{i1,}, lb_{i2}, ...lb_{iD}) = (x_{i1}, x_{i2}, ...x_{iD})$

Similar is the case with gbest and Itbest.

It has been found that velocity bounds, inertia weights and learning rates have a direct effect on the particle's motion[25]. Hence, it is important to assign values to these three parameters in such a manner that it can effectively control the diversification and intensification of the particle. The values for these respective parameters have been set based on experimentation as well as the settings used in the previous papers [26].

- The velocity bounds include the upper velocity bound ($V_{\max}$) and lower velocity bound($V_{\min}$). They define the maximum and minimum velocity values that any velocity V_{ij} can take i.e.

  if $V_{ij} > V_{\max}$ then $V_{ij} = V_{\max}$

  if $V_{ij} < V_{\min}$ then $V_{ij} = V_{\min}$

  Here $V_{\max}$=6 and $V_{\min}$=-6
- Inertia weight ω is updated using the following expression

$$\omega = \omega_{\max} - \frac{(\omega_{\max} - \omega_{\min})t}{T} \tag{6}$$

  where $\omega_{\max}$ and $\omega_{\min}$ are the upper and lower bounds for inertia weight, t is the current iteration and T is the total number of BPSO iterations Here $\omega_{\max}$=0.995, $\omega_{\min}$=0.5,T=100.
- The learning rates c_1, c_2 are set as 1.49618 and c_3 as0.5
- The number of the particles or the solutions is initialized as 30 i.e. N=30. The position of the particle is defined as the binary vector X_i={ $x_{i1,}, x_{i2}, x_{i3}$.....x_{iD} } where D represents the total number of features i.e. D=117. $x_{ij} \in \{0,1\}$, 1 if the feature is selected, 0 otherwise. d represents the total number of optimal features. d is initialized as 20. i.e. $\sum_j x_{ij} = d$. The initial velocity of any particle i is taken as zero.

4.1 k-Nearest Neighbor with Leave One Out Cross Validation (kNN-LOOCV)

The choice of fitness function for evaluating the quality of features selected is an important decision in PSO based feature selection methods. One of the popular fitness functions is the classification accuracy of the induced model. In the proposed methodology, k-nearest neighbor classification accuracy using leave-one–out cross validation has been used as the fitness function. This means that modified BPSO searches for the optimal feature subset and k-NN classifier evaluates each feature subset based on its classification accuracy using leave-one-out cross validation.

k-NN algorithm[27] is a simple and popular non parametric method which stores a training dataset of instances and classifies a query instance based on the attributes and similarity measure of the instance with that of the training set. The instance is assigned a class most common among its k -closest neighbours in the training set as measured by the Euclidean distance. The accuracy of the classification is measured using leave-one-out cross validation (LOOCV) [28]. Suppose thetraining data set consists of n instances. At each iteration, LOOCV uses one instance from the training data set as test data and the remaining n-1 instances as the training data. k -NN classifier

is applied to find the class of the test data. This procedure is repeated for all the remaining instances. Accuracy of the classifier is calculated as the ratio of the total number of correctly classified instances to that of the total number of instances.

4.2 Modified Binary Particle Swarm Optimization Algorithm

Input:Training data set.

Output: Selected feature subset G

1. Initialize $c_1 = c_2 = 1.49618, c_3 = 0.5, V_{\max} = 6, V_{\min} = -6, \omega_{\max} = 0.995, \omega_{\min} = 0.5,$ D=117, d=20,T=100, t=0, n=30.
2. Randomly generate n initial particles which are binary vectors of length D such that the total number of binary ones in each vector is d. i.e. $\sum_j x_{ij} = d.$

```
for i=1 to n;
    for j=1 to D
        x_ij =1 or 0
    next j
next i
```

3. Initialize the velocity of n particles as 0.

```
for i=1 to n;
    for j=1 to D
        v_ij = 0
    next j
next i
```

4. Calculate lbest of each particle, Itbest and gbest

4.1 for each particle i=1 to n.

- Calculate the fitness function using k-NN LOOCV $lbest_i$=K-NN-LOO(X_i).
- Update the position vector of $lbest_i$ with the position of the particle.

$(lb_{i1,}, lb_{i2}, ...lb_{iD}) = (x_{i1}, x_{i2}, ...x_{iD})$

next i.

4.2 Update Itbest with the highest fitness value obtained among particles in the current iteration.

4.3 Assign the position vector of the corresponding particle to that of the Itbest.

(It_1, It_2…… It_D)= $(x_{i1}, x_{i2}, ...x_{iD})$

4.4 Update gbest with the highest fitness value obtained among all particles so far.

4.5 Assign the position vector of the corresponding particle to the position vector of gbest.

(gb_1, $gb_{2....}gb_D$)= $(x_{i1}, x_{i2}, ...x_{iD})$

5. Repeat while (t≤T || gbest <= 0.99)

t=t+1;

Update ω using equation(6).

Generate a random no δ between 0 and 1

5.1 for i=1 to n

for j=1 to D

- Calculate velocity v_{ij} using equation (5)
- Update x_{ij} using the sigmoid function given in equation(3) and (4).

next j.

next i.

5.2 Calculate the new fitness value for particle X_i using K-NN-LOO().

fitness (X_i)=K-NN-LOO(X_i)

5.3 Update $lbest_i$ if the following condition holds true.

If((fitness(X_i)>$lbest_i$) ‖ ((fitness(X_i)=$lbest_i$)&&($|X_i| < |lbest_i|$)))

$lbest_i$= fitness(X_i)

($lb_{i1}, lb_{i2} \ldots lb_{iD}$)= ($x_{i1}, x_{i2} \ldots x_{iD}$) //Update position vector

end if.

5.4 Assign Itbest with the best local best value obtained in the current iteration t.

If (($lbest_i(t)$ > Itbest) ‖ (($lbest_i(t)$ = Itbest)&&($|lbest_i(t)| < |ltbest|$)))

Itbest= $lbest_i$

($It_1, It_2 \ldots It_D$)= ($x_{i1}, x_{i2} \ldots x_{iD}$)

end if

5.5 Update gbest with the best local best value obtained so far.

If (($lbest_i$>gbest) ‖ (($lbest_i$= gbest)&&($|lbest_i| < |gbest|$)))

gbest= $lbest_i$

($gb_1, gb_2 \ldots gb_D$)= ($x_{i1}, x_{i2} \ldots x_{iD}$)

end if

end Repeat

Return selected feature subset G where jϵG if gb_j=1

End

Procedure K-NN-LOO(X_i)

Begin

1. for j= 1 to m // m is the number of objects in training set.

- Temporarily remove j^{th} object(O_j) from the training set .
- Euclidean distance between the j^{th} object (O_j) to all the remaining (m-1)objects in the training set is found. To calculate the Euclidean distance only the features corresponding to binary bit 1 in the position vector of particle X_i is considered.

fork=1 to D

If(x_{ik} =1)

Dist(O_j, O_l)= Sqrt(sum+($O_{jk} - O_{lk}$)^2) O_l ϵ training set.

end if

next k.

- Find the K nearest neighbors to O_j which has the minimum Euclidean distance.

- The most common class among K neighbors is assigned to O_j .

next j.
2. for j=1 to m

- If (Class(O_j)= real class of (O_j))

correct = correct + 1;
end if.
next j.
3. fitness value= correct/|training set|.
return(fitness value)
end

5. EXPERIMENTAL RESULTS

This section describes the database used, the test methodology and the results obtained, and comparison of the proposed feature selection method MBPSO with other existing techniques.

5.1 Database

In order to evaluate the performance of the modified BPSO for feature selection, digital mammograms have been taken from 2 sources. 83 mammograms have been provided by the Regional Cancer Center, Thiruvananthapuram. All images are from Hologic Selenia Dimensions full field digital mammograms system installed at Regional Cancer Center. These images are in DICOM format with a resolution of 4096×3328 pixels. They have a pixel size of 65µm and bit depth of 12 bits. 32 mammograms are malignant and the remaining 51 are normal.

200 mammograms have been taken from mini-MIAS database [29] which is a web- accessible international resource. All images are in portable gray map format(.pgm). They are digitized at a partial resolution of 0.05 mm pixel size with 2 bit density resolution using SCANDIG-3. All have been expertly diagnosed and positions of the abnormalities have been recorded. 127 mammograms are normal, 44 are benign and 29 of them are malignant. For this experiment, a total of 283 mammograms are used of which 178 are normal, 44benign and 61 malignant.

5.2 Implementation Environment

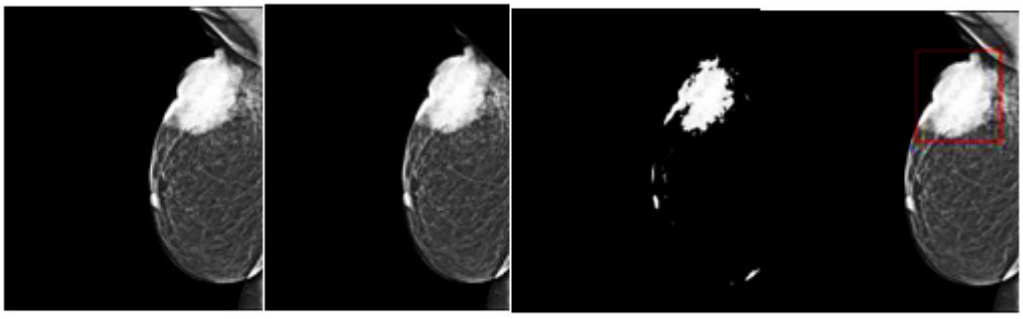
The experiment is implemented on Windows 10 Pro 64-bit operating system using MATLAB 2015b 64-bit, with Matlab image processing tools and statistical tools. All experiments are implemented on Intel Core x64-based Processor of 2.4 GHz CPU with 8GB RAM.

5.3 Test methodology

All the digital mammograms have been preprocessed, ROIs are automatically segmented and features are extracted as shown in our previous work [30]. Image preprocessing is required to enhance the breast profile and to remove artefacts, labels, noise that can appear accidentally in mammograms. It is also required to remove the unrelated parts that may appear in mammograms like pectoral muscles. Median filter, global thresholding, adaptive fuzzy logic based bi-histogram equalization [31] has been used to remove labels, artefacts and to obtain controlled enhancement. In order to remove pectoral muscles, bounding box of the image has been used. Suspicious space occupying lesions are automatically segmented from the mammograms for further preprocessing. Multithresholding based on Otsu's method and morphological operations are used for segmentation of the ROIs. Normal mammograms do not contain lesions. But, same procedure for

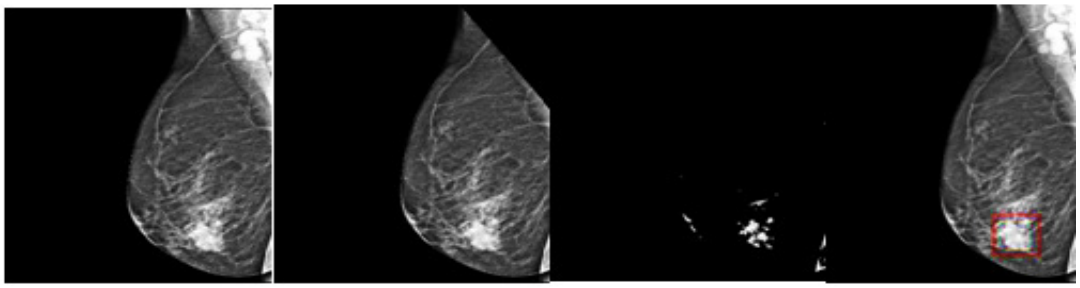
pre-processing and segmentation are applied to them as well and ROIs are extracted. Figures 1 and 2 show the pre-processing and segmentation results obtained when applied on two mammograms. Shape, texture and grey level intensity values of the ROI play an important role in differentiating them as healthy, benign or malignant [32]. Therefore, 6 grey level intensity features (mean, variance, skewness, kurtosis, energy and entropy), 52 GLCM features (energy, contrast, correlation, variance, homogeneity, entropy, sum average, sum entropy, sum variance, difference variance, first correlation measure and second correlation measure in four directions 0^o,45^o,90^o,135^o)44GLRLM features (SRE,LRE,GLN,RP,RLN,LGRE, HGRE,SRLGE,SRHGE, LRLGE and LRHGE in four directions 0^o,45^o,90^o, 135^o) and 15 shape features (area, perimeter, eccentricity, equidiameter, compactness, Thinness Ratio, Circularity, elongatedness, dispersion, Shape index, Euler number, SD of edge and mass, Max Radius and Min Radius) are extracted. Hence, a total of 117 features are extracted from the ROIs. These features are taken as input for the feature selection technique.

ROIs obtained are divided into 2 sets. One set is used as the training set and the other set as the test set. Training set contains 227 ROIs (144 normal, 35 benign and 48 malignant).There are 56ROIs in the test set (34 normal, 9 benign and 13 malignant). Feature selection using MBPSOis done using the training set only. This results in an optimal feature set of 6 features. Theoptimal features obtained are used to train the classifier, using the training set. A feed forward artificial neural network with back propagation (FFANN) [33] has been used as the classifier for the classification phase. It consists of three layers- an input layer with number of neurons equal to number of selected features, a hidden layer made up of neurons and an output layer with three neurons each representing a target class- normal, benign and malignant. Initial weights and bias are randomly selected for FFANN usually between -0.1 to 1.0 and -0.5 to 0.5. To propagate the inputs forward, non-linear log sigmoid function is used as the activation function. A matrix of size 6×227 is given as input to the input layer. Using this feature matrix, FFANN processes the data by comparing the network prediction of each tuple with the actual known class label. FFANN learns using gradient descent method in the backward direction to iteratively search for a set of weights and bias to minimize the mean-squared distance between network's class prediction and the known target value of the tuples. After the necessary accuracy is obtained, the weights are frozen. The test data is then fed to the FFANN. For each test mammogram, a column vector is created where each element represents one of the optimal features for the respective mammogram. The class is then decided by FFANN based on the training results.



(a) (b) (c) (d)

Figure1. (a) Mammogram image PAT0004 obtained from Regional Cancer Center, Thiruvananthapuram. (b) PAT0004 with pectoral muscles removed. (c) ROI obtained after segmentation. (d) Marked malignant portion.



(a) (b) (c) (d)

Figure 2. (a) Mammogram image PAT0014 obtained from Regional Cancer Center, Thiruvananthapuram (b) PAT0014 with pectoral muscles removed (c) ROI obtained after segmentation (d) marked malignant portion.

The MBPSO feature selection method is compared to three other feature selection methods-PCA, RFE and CART. The three methods are briefly described below.

- Principle Component Analysis (PCA): PCA [34] is a linear transformation method to compress the data by reducing the number of dimensions without loss of information. It makes use of covariance matrix, Eigen vectors and Eigen values to find the components and then forms the optimal feature subset based on the components that are chosen.

- Recursive feature elimination (RFE): RFE [35] is a wrapper feature selection method which works by recursively removing weaker attributes and building a model based on those attributes that remain. It uses a model accuracy to identify which attributes contribute the most to predicting the target. The stability of RFE depends on the type of model that is used for feature ranking. The model used here is support vector machine.

- Classification and Regression Tree (CART): CART [36] is a decision tree induction algorithm which constructs a flow chart like attribute structure where each internal node denotes a test attribute and each external node denotes a class prediction. Since at each node, the algorithm choses the best attribute to partition the data into individual classes, these attributes can be taken as significant features and they form the reduced subset of features.

In order to maintain uniformity, the same training set is used by all feature selection methods to find the optimal features. Using the optimal features, training dataset is used to train the FFANN. The test data set is then fed into the classifier and classification accuracy is measured.

5.4 Experimental Studies

In this section, a series of experiments are carried out to evaluate the accuracy and efficiency of the proposed method MBPSO. As mentioned before, the same training data set is used for all feature selection methods to find the optimal features and the same test data is used to measure the classification accuracy. Classification accuracy is defined as the total number of correctly classified samples in the test data divided by the total number of test samples.

Each feature selection method is run 3 times and the optimal feature subset which provides the best classification accuracy is chosen. In case, two feature subsets have the same accuracy, the one with lesser number of features is chosen. Table 1 gives the comparison of various feature selection methods based on their classification accuracy and computational time.

Table 1. Comparison of various feature selection methods based on their classification accuracy and computational time.

Feature selection methods	Number of optimal features	Classification accuracy	Computational time
All features	117	87.3	0.610460
PCA	6	93.6	0.571530
CART	9	96.5	0.582645
RFE	11	92.9	0.596442
Proposed method	6	97.2	0.583761

Figures 3 -7 demonstrate the classification performance of various feature selection methods using ROC curve and all confusion matrices.Fig 3 represents all confusion matrix obtained when no feature selection is used. Out of 178 normal cases, 44 benign cases and 61 malignant cases, 172 normal cases, 30 benign cases and 45 malignant cases have been correctly classified giving a classification accuracy of 87.3%.Fig 4 shows that the classification accuracy obtained when PCA is used as the feature selection method is 93.6%.Fig 5 and Fig 6 demonstrates that the classification accuracy obtained is 96.5% and 92.9% respectively when CART and RFE are used as feature selection methods.Fig 7 shows the classification accuracy obtained when the proposed method MBPSO is used as the feature selection method. It obtains a classification accuracy of 97.2% as 177 normal,42 benign and 56 malignant cases have been correctly classified.

From Table 1 and Figure 7, it can be seen that MBPSO reduces the feature set from 117 to 6. The optimal features include SD of the edge, entropy, LGRE(0^0), contrast (90^0), perimeter, SGLGE (0^0).It gives a classification accuracy of 97.2% which is better than the classification accuracy without feature selection.

PCA though reduces the feature subset to 6, the classification accuracy obtained is93.6% which is less when compared to CART and the proposed method. Studies have shown thatPCA is unable to capture accurately nonlinear relationships which exists in the complex biological systems.This may be the reason for the reduced accuracy. The advantage of PCA is that PCA takes lesser computational time when compared to other feature selection methods.

CART provides an optimal subset of 9 features and results in a classification accuracy of 96.5%. The 9 features include LGRE (135^0), HGRE(90^0), LRLGE (90^0), SD of Edge, LRLGE (0^0), LRE (0^0), LGRE (90^0), LGRE(0^0), contrast (90^0).The computational time is also almost similar to that of MBPSO. But the advantage of MBPSO over CART is that it uses lesser number of features as can be seen from Table 1, figure 7 and figure 5. MBPSO also attains a sensitivity of 91.8% and specificity of 95.5% in classification whereas CART attains a sensitivity of 88.4% and specificity of95.4% .This means MBPSO results in greater number of malignant and benign cases being correctly classified as compared to CART. This is due to the fact that CART can take only one attribute at a time to make the split (decision) and thus if the decision depends on several variables, chances of error rate is higher.To differentiate between normal, benign and malignant breast tissues, shape, texture and histogram features have to be considered simultaneously in order to make correct diagnosis. As this is not possible in case of CART, this may have resulted in lesser sensitivity and specificity.

SVM-RFE provides an optimal subset of 11 features. They include entropy, SD of edge, homogeneity, LGRE (90^0), LGRE(0^0), contrast (90^0), difference entropy (135^0), difference variance (135^0), SRE (45^0). However, RFE is computationally expensive as compared to MBPSO and other feature selection methods. This is because SVM-RFE goes through each feature one by

one in order to remove weaker attributes and to build a model based on the optimal attributes. It also does not take into account the correlation between features.

The experimental results prove the efficacy of the proposed method and also show that the performance of the proposed method is at par with other popular feature selection methods.

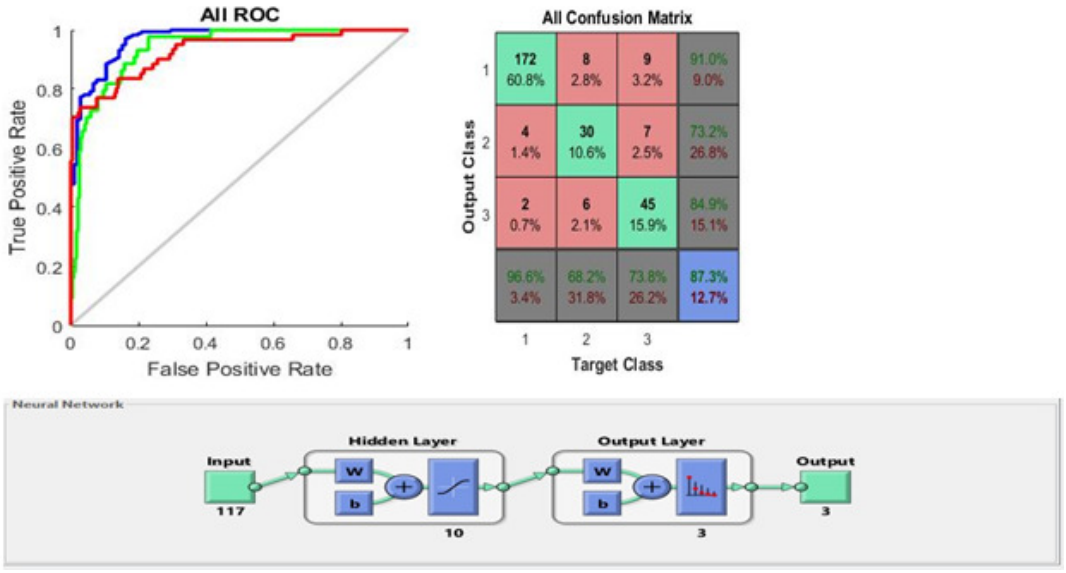


Figure 3. Classification performance without feature selection

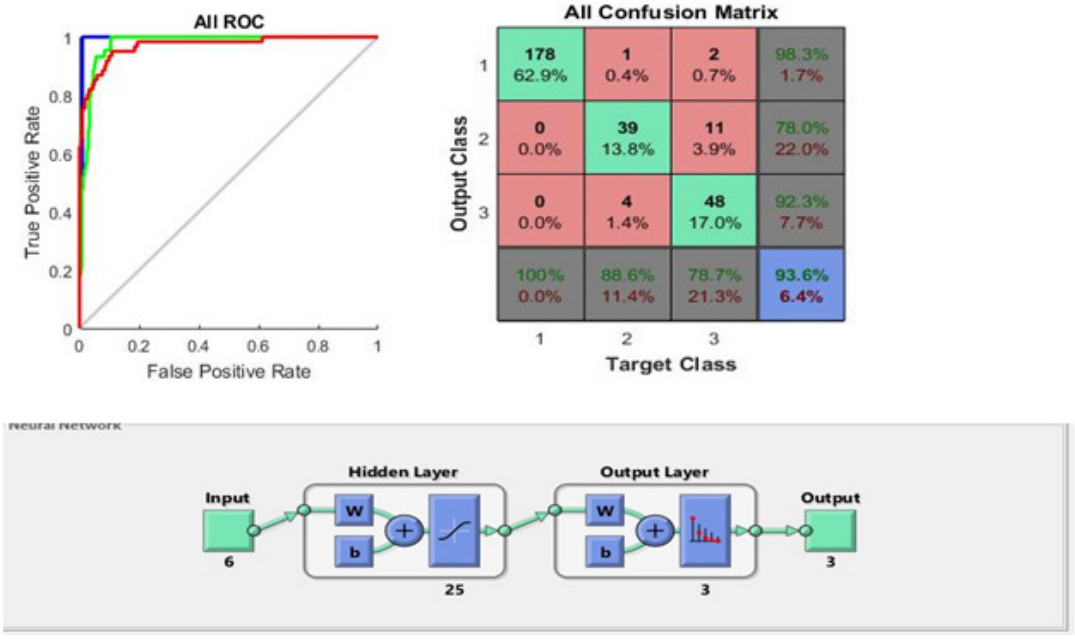


Figure4. Classification performance with PCA as feature selection method.

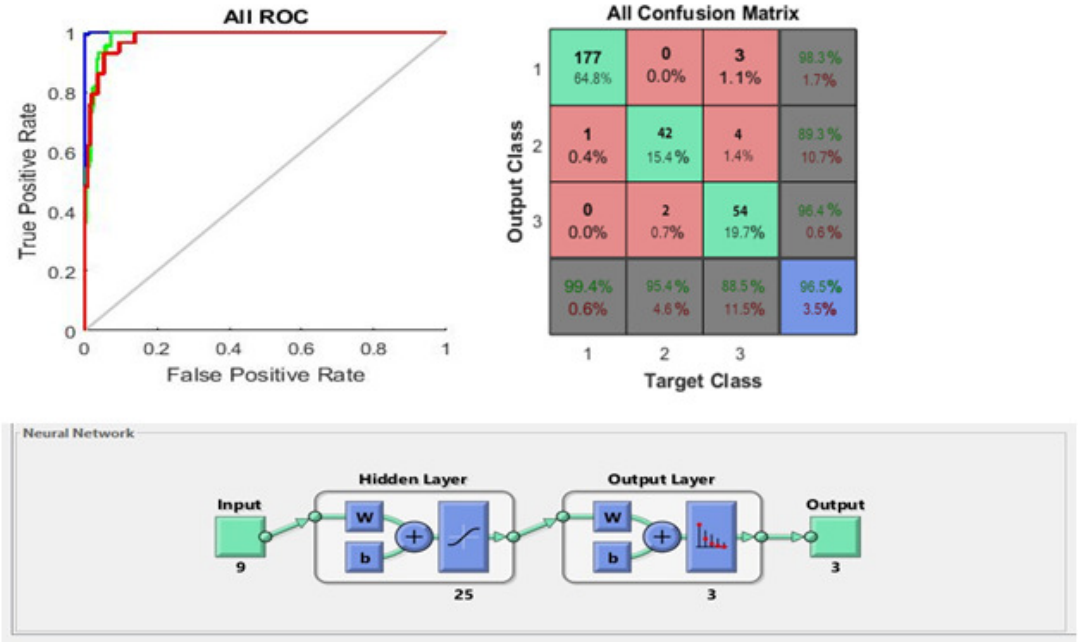


Figure 5 Classification performance with CART as feature selection method.

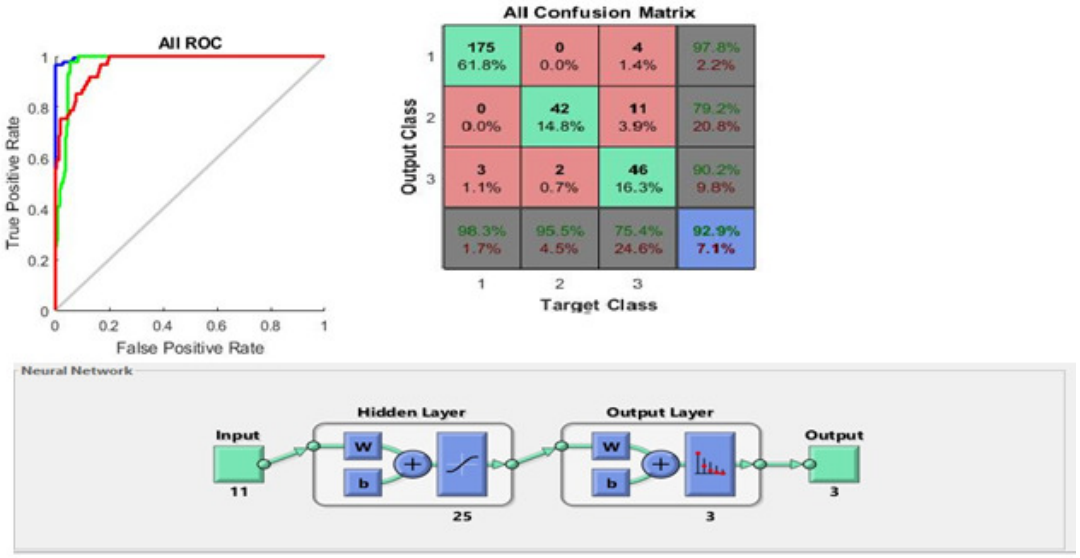


Figure 6. Classification performance with RFE as feature selection method.

Figure 7. Classification performance with the proposed method as feature selection method.

6. CONCLUSION

Feature selection is an important pre-processing tool in building an efficient classification model. In this paper, a modified binary PSO-KNN method for feature selection has been proposed in order to develop a classification model to distinguish between healthy, benign and malignant parenchyma in mammograms. Experimental results show that the proposed method obtainsan accuracy comparable to other popular feature selection methods. At the same time, it reduces computational complexity and also demonstrates high efficiency which is atpar with other well-known feature selection methods. In future, modified BPSO can be applied to problems in other areas as well.

REFERENCES

[1] R Siegel, D. Naishadham and A.Jeimal(2013)"Cancer Statistics 2013", CA: A Cancer Journal for Clinicians, Vol. 63, pp. 11-30.

[2] (2012-2014) Statistics of Breast Cancer in India. "Trends of Breast Cancer in India" [online] Availablehttp://www.breastcancerindia.net/statistics/trends.html

[3] (2013-14) World Cancer Research Fund International, "Cancer Facts and Figures" [online] Available http://www.wcrf.org/int/cancer-facts-and-figures

[4] (2016) American Cancer Society, "Breast Cancer Signs and Symptoms" [online] Available https://www.cancer.org/cancer/breast-cancer/about/breast-cancer-signs-and-symptoms.html

[5] K.Hu, X. Gao and F.Li (2011) "Detection of suspicious lesions by adaptive thresholding based on multi resolution analysis in mammograms", IEEE Transactions on Instrumentation and Measurement, Vol. 60 (2), pp. 462-472.

[6] K.U Sheba and S. Gladston Raj (2016) "Objective quality assessment of image enhancement methods in digital mammography-A comparative study",Signal and Image Processing: An International Journal, Vol. 7(4), pp.1-13.

[7] M.J.G Calas, B Gutfilen and W.C.A Pereira (2012) CAD and Mammography: Why use this tool?", Radilogical Brasileira, Vol. 45 (1) , pp. 46-52.

[8] J.Deeba and S.T Selvi(2014) "Computer-aided detection of breast cancer on mammograms: A Swarm Intelligence optimized wavelet neural network approach", Journal of Biomedical Informatics, Vol. 49, pp.45-52.

[9] El-Baz, G.M Beache, G Gimel'farb et.al., (2013)" Computer-aided diagnosis system for lung cancer: challenges and methodologies", International Journal of Biomedical Imaging, Vol 2013, Article T D 942353, 46 pages.doi:10,1155/2013/942353.

[10] K.U Sheba and Gladston Raj S.,(2017) "Detection of lesions in mammograms using grey-level, texture and shape features", Journal of Advanced Research in Dynamical and Control Systems, Vol. 9 Sp-16, pp. 919-936.

[11] B.I Shak and Anis(2016) "Variable selection using support vector regression and random forests: A comparative study",Intelligent Data Analysis, Vol. 20 (1), pp. 83-104.

[12] B. Xue, MZhang and W.N Browne(2012) "New fitness functions in binary particle swarm optimization for feature selection", IEEE World Congress on Computational Intelligence (WCCI 2012),Brisbane, Australia.

[13] A.Unler and A. Murat(2010) "A discrete particle swarm optimization method for feature selection in binary classification problems", European Journal of Operational Research, Vol. 206, pp. 528-539.

[14] V. Kothari, J. Anuradha, S. Shah and P.Mittal (2012) "A survey on particle swarm optimization in feature selection", In: Krishna P.Y, Babu M.R, Ariwa E. (eds), Global Trends in Information systems and software applications, Vol. 270, pp. 192-201.

[15] X. Wang, J. Yang, X. Teng, W. Xia and R. Jensen(2007)" Feature selection based on rough sets and particle swarm optimization", Pattern Recognition Letters, Vol. 28(4), pp. 459-471.

[16] B.Xue, S. Nguyen and M. Zhang(2014) "Anew binary particle swarm optimization algorithm for feature selection", In: Esparcia-Alca'zar A., Mora A.(eds). Applications of Evolutionary Computation. EvoAapplications 2014, Lecture Notes in Computer science, Vol. 8602. Springer Berlin, Heidelberg. pp 501-513.

[17] B.Tran, B. Xue and M. Zhang (2014) "Overview of PSO for feature selection in classification", In: Dick G et.al (eds). Simulated Evolution and Learning (SEAL 2014),Lecture Notes in Computer Science, Vol. 8886, Springer, Cham, pp. 605-617.

[18] Z.Yong, G. Dunwei, H.Ying and Z. Wanqiu (2015) "Feature selection algorithm based on bare bones PSO". Neurocomputing, Vol.148(6), pp. 150-157.

[19] M.T Wong, X.He, W.C. Yeh, Z Ibrahim and Y.Y Chung (2014) "Feature selection and mass classification using PSO and SVM", In: Loo C.K, Yap K.S, Wong K.W, Beng Jin A.T, Huang K (eds)., Neural information processing. ICONN 2014, Lecture Notes in Computer Science, Vol. 8836, Springer, Cham, pp. 439-446.

[20] S. Sivakumar and C. Chandrasekhar (2014) "Modified PSO based feature selection for classification of lung CT images", International Journal of Computer Science and Information Technologies, Vol. 5(2), pp. 2095-2098.

[21] I. Zyout and I. Abdel-Qader(2011) "Classification of microcalcification clusters via PSO-KNN heuristic parameter selection and GLCM features", International Journal of Computer Applications, Vol. 31(2), pp. 34-39.

[22] M.T Wong, X.He and H Nguyen (2012) "Particle Swarm Optimization based feature selection in mammogram mass classification", Computerized Health Care (ICCH 2012), International conference on, pp.152-157, Dec 2012.

[23] J. Kennedy and R. Eberhart(1995) "Particle Swarm Optimization", In:Proceedings of the 1995 IEEE international conference on neural networks,Perth, Australia, Vol. 4,pp. 1942-1948.

[24] J. Kennedy and R. Eberhart (1997), "A discrete binary version of particle swarm algorithm", In: Proceedings of the 1997 IEEE International Conference on Systems, Man and Cybernetics (SMC 97), Vol. 5, pp. 4104-4108.

[25] Y.Shi and R. Eberhart(1998) "A Modified Particle Swarm Optimizer", In: Proceedings of IEEE International Conference on Evolutionary Computation, World Congress on Computational Intelligence, Anchorage, Alaska.

[26] B.Xue, M.Zhang and W.N Browne(2012) " New Fitness Functions in binary particle swarm optimization for feature selection, IEEE World Congress on Computational Intelligence (WCCI 2012),Brisbane, Australia.

[27] S.Zhang, X.Li, M.Zong, X. Zhu and R.Wang (2017) "Efficient KNN classification with different numbers of nearest neighbors", IEEE Transactions on Neural Networks and Learning systems, Vol. 99, pp.1-12.

[28] A.Vehtari, A.Gelman andJ.Gabry (2017) "Practicalbayesian model evaluation using leave-one-out cross validation and WAIC", Statistics and Computing, Vol. 27(5),pp.1413-1432.

[29] J Suckling et.al (1994). The mammographic Image Analysis Society Digital Mammogram database. Exerpta Medica. International Congress Series 1069, pp. 375-378.

[30] K.U Sheba and S. Gladston Raj (2017) "Detection of Lesions in Mammograms using grey-level, texture and shape features", Journal of Advanced Research in Dynamical and Control Systems, Vol.(16)-Special Issue, pp. 919-936.

[31] K.U Sheba and S. Gladston Raj (2017) "Adaptive fuzzy logic based bi-histogram equalization for contrast enhancement of mammograms", In: Proceedings of IEEE International Conference on Intelligent Computing, Instrumentation and Control Technologies, Kannur, Kerala.(to be published)

[32] M.J Homer (2004) "Breast Imaging, Standard of care and the expert",Radiologic Clinics of North America, Vol. 42(5), pp. 963-974.

[33] P.Tahmasebi and A.Hezarkhani (2011) "Application of a modular feed forward neural network for grade estimation", Natural Resources Research, Vol. 20(1), pp. 25-32.

[34] Z.M Hera and D.F Gillies (2015) " A review of feature selection and feature extraction methods applied on micro array data", Advances on Bioinformatics, Vol. 2015, Article ID 198363, 13 pages. doi 10.1155/2015/198363

[35] I. Guyon, J. Weston, S.Barnhill and V.Vapnik, (2002) "Gene selection for cancer classification using support vector machines", Mach.Learn, Vol. 46(1-3), pp. 389-422.

[36] T.Hayes, S.Usami et.al, (2015) "Using classification and regression Trees (CART) and random forest to analyze attrition: Results from two simulations", Psychology Aging, Vol. 30(4), pp. 911-929.