

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/323610902>

Classification of Masses in Digital Mammograms Using Firefly based Optimization

Article in International Journal of Image, Graphics and Signal Processing · February 2018

DOI: 10.5815/ijigsp.2018.02.03

CITATIONS

4

READS

23

2 authors, including:



[Shankar Thawkar](#)

Hindustan College of Science and Technology

7 PUBLICATIONS 24 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Digital Mammography [View project](#)

Classification of Masses in Digital Mammograms Using Firefly based Optimization

Shankar Thawkar

Department of Information Technology, Hindustan College of Science and Technology,
Farah Mathura, 281122, India
Email: Shankar.thawkar@gmail.com

Ranjana Ingolikar

Department of Computer Science, S. F. S. College
Nagpur, 440001, India
Email: r_ingolikar@rediffmail.com

Received: 21 August 2017; Accepted: 17 November 2017; Published: 08 February 2018

Abstract—Breast cancer is one of the leading causes of death in women all over the world. Computer based diagnosis system assists radiologist in the effective treatment of breast cancer. To design an efficient classification system for masses in digital mammograms, we have to use efficient algorithms for feature selection to reduce the feature space of mammogram classification problem. The proposed study explores the use of Firefly algorithm to select a subset of features. Artificial neural network and support vector machine classifiers are employed to evaluate fitness of the selected features. Features selected by Firefly algorithm are used to classify masses into benign and malignant, using artificial neural network and support vector machine classifiers. The proposed method employed over 651 mammograms obtained from the Database of Digitized Screen-film Mammograms. Classification results show that Firefly algorithm with artificial neural network is superior to Firefly algorithm with support vector machine. Artificial neural network achieves accuracy of 95.23% with 94.43% sensitivity, 93.94% specificity and area under curve $Az=0.965\pm0.008$. On the other hand, support vector machine classifier achieves an accuracy of 92.47% with 96.14% sensitivity, 88.53% specificity and area under curve $Az=0.951\pm0.009$. Results obtained with Firefly algorithm shows that it will be useful for effective treatment of breast cancer.

Index Terms—Firefly algorithm, artificial neural network, support vector machine, receiver operating characteristics curve, digital mammography, feature selection, classification.

I. INTRODUCTION

Breast cancer is one of the leading causes of death in women all over the world. Currently no technique is available for prevention of breast cancer; hence, detection of breast cancer in its primary stage is very important. Mammography is among the best available techniques for

detection of breast cancer in the primary stage [1-2]. In this technique, each breast image, called mammogram, is captured from two views called Craniocaudal (CC) and Mediolateral-oblique (MLO). Radiologist diagnoses breast cancer by reading mammograms, though reading of mammograms is a very challenging task. Therefore, for clinical observation, suspicious masses have to be removed from the breast using biopsy procedure. Available facts show that more than sixty to seventy percent of suspicious cases turn out to be benign. This problem of unnecessary biopsies can be minimized with the use of computer based diagnosis system. These systems function as a second opinion for radiologists, and help to improve breast cancer diagnosis. Advancements in the area of image processing and machine learning have made detection and classification of masses in digital mammograms easier, but it remains a challenging area of research. The proposed study focuses on the classification of masses in digital mammograms. Available literature shows that the performance of classifiers depends on feature selection. Feature selection method is intended to remove irrelevant or redundant features that may hamper performance of the classifier, from the extracted feature set.

In digital mammography, automatic mass detection and classification are challenging fields of research. In the past, many techniques were developed for feature optimization based on swarm intelligence. These techniques have evolved from the field of artificial intelligence [3]. Some of the swarm intelligence based techniques are: Artificial Bee Colony (ABC) [4], Ant Colony Optimization (ACO) [5] and Particle Swarm Optimization (PSO) [6]. The most promising swarm intelligence based techniques used currently are Firefly Algorithm (FA) [7-8], cuckoo search [9], bat algorithm [10] and kill herd bio-inspired optimization algorithm [11].

Firefly Algorithm is a classification technique developed recently. It is a nature inspired meta-heuristic algorithm developed by Xin-She Yang in 2008 and 2009

[7-8], based on the flashing pattern of fireflies. It has been employed in several applications such as, classification, clustering, optimization and NP-hard scheduling problems [12].

The rest of the paper organized as: Section II presents Related work. Proposed framework is presented in Section III. Section IV presents results of different methods and discussion about their performances. Conclusion of the paper is given in Section V.

II. RELATED WORK

Saraç et al. [13] studied Firefly algorithm to select features for web page classification. FA selects feature subsets based on fitness of the selected feature. J48 classifier is used to find fitness of the selected feature. Senthilnath et al. [14] used Firefly algorithm for clustering on a benchmark data set obtained from UCI machine learning data repository. The performance of FA was comparable with ABC and PSO methods. Results show that FA is superior to both ABC and PSO methods. Banati et al. [15] investigated the use of rough set theory with firefly algorithm for feature selection. Performance of the method was tested on a medical data set obtained from UCI machine learning data repository [16]. The method proves to be best in terms of time and optimality as compared to other methods.

Pal et al. [17] perform the comparative study between Firefly algorithm and Particle Swarm optimization (PSO) for finding the optimal solution of noisy non-linear optimization problems. The outcomes of the method show that, FA is better than PSO for higher level of noise. Singh et al. [18] study the use of Krill Herd algorithm (KH), Firefly algorithm (FA) and Cuckoo search Algorithm (CS) for finding optimal solutions of various unimodal and multimodal mathematical test functions. The performances of these techniques were tested based on efficiency, convergence and time. The results show that Firefly algorithm outperforms Krill Herd for unimodal optimization.

Dhal et al. [19] investigate the use of Bat algorithm (BA) and Firefly algorithm for image enhancement. These techniques are used to optimize the parameters of high boost filter. The goodness of image enhancement is tested using objective function. The image parameters, Entropy and number of edge pixels act as an objective function.

In many applications, performance of Firefly algorithm is found superior to other swarm intelligence based algorithms. This motivates us to employ it to select significant features for the classification of masses in digital mammograms into benign and malignant.

III. PROPOSED FRAMEWORK

The proposed framework consists of two main steps. In the first, we used Firefly algorithm for feature selection, and to select optimal features, artificial neural network (ANN) and Support Vector Machine (SVM) classifiers

were used. In the second step, features selected were used to classify suspicious masses into benign and malignant using, ANN and SVM. Overview of proposed framework is as shown in Fig. 1.

A. Feature Extraction

The process of automatic mass detection and feature extraction has been discussed in a previous work [20]. The features extracted from each of the detected masses are divided into three types, intensity based (FNo. 1 to 6), texture based (FNo. 7 to 17) and shape based (FNo. 18 to 25). The list of features extracted from mammograms is shown in Table 1 [20-21].

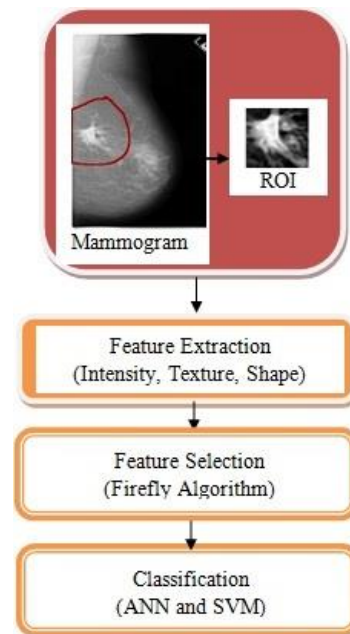


Fig.1. Proposed framework

Table 1. List of extracted features

FNo.	Features	FNo.	Features
1	Average gray level	14	Homogeneity
2	Average contrast	15	Sum average
3	Smoothness	16	Sum Variance
4	Skewness	17	Sum entropy
5	Uniformity	18	Area
6	Entropy1	19	Perimeter
7	Energy	20	Compactness
8	Entropy2	21	Normalized Standard deviation
9	Contrast	22	Area Ratio
10	Mean	23	Contour Roughness
11	Standard deviation	24	Normalized Residual Value
12	Variance	25	Overlapping Ratio
13	Correlation		

B. Feature Selection using Firefly Algorithm

Basic purpose of the feature selection technique is to remove irrelevant or unnecessary features from the extracted feature set. It selects the most relevant features

that would improve the performance of the classifier. Feature selection techniques select the most relevant features based on four criteria: Discrimination, Reliability, Independence and Optimality [22-23]. In this article, Firefly Algorithm (FA) is used to select a subset of features. Artificial Neural Network (ANN) and Support Vector Machine (SVM) classifiers are employed to evaluate fitness of the selected features.

Firefly Algorithm (FA) was developed by Xin-She Yang in 2008 and 2009 [7-8] at Cambridge University, inspired by the flashing patterns and idealized behavior of fireflies. Flash patterns produced by the process of bioluminescence are unique for each firefly species. More than two thousand firefly species exist worldwide. Primary function of the flashing lights is to attract other fireflies for mating and also to attract potential prey. The flashing lights also serve as a defense mechanism for fireflies to warn potential predators. FA uses the following three idealized rules:

- i. All fireflies are unisex.
- ii. Attractiveness is proportional to brightness. So, considering two flashing fireflies, the less bright one will move towards the brighter one. Both attractiveness and brightness decrease as distance increases. A firefly moves randomly if there is no brighter firefly than itself.
- iii. Light intensity (brightness) of a firefly is affected or determined by the landscape of the objective function.

In maximization problem, the objective function $f(z)$ is similar to fitness function used in genetic algorithms and it is proportional to brightness $I(d)$. That is $I(z) \propto f(z)$, for a particular location z . The attractive parameter β , differs with distance r_{ij} between firefly 'i' and 'j'. The attractiveness parameter also varies with absorption in the medium.

Intensity of light, $I(r)$, in the maximization problem, is formulated as –

$$I(r) = \frac{I_s}{r^2} \quad (1)$$

where I_s , is the intensity at light source and r is the distance between fireflies.

The variation of light intensity I with respect to distance r for a specified medium with fixed light absorption coefficient γ , is defined as –

$$I = I_0 e^{-\gamma r} \quad (2)$$

where, I_0 is the initial light intensity. The combined effect of inverse square law and absorption is approximated using Gaussian form as defined in (3). It is used to avoid singularity at $r=0$ in I_s/r^2 .

$$I(r) = I_0 e^{-\gamma r^2} \quad (3)$$

The other form of function $I(r)$, is defined as-

$$I(r) = \frac{I_0}{1+\gamma r^2} \quad (4)$$

Equation (4) is implemented, when $I(r)$ is required to be decreased monotonically at a slower rate. The functions defined in Equations (3)-(4) are essentially the same. Attractiveness parameter β is proportional to the light intensity seen by the neighboring firefly, defined as

$$\beta(r) = \beta_0 e^{-\gamma r^2} \quad (5)$$

where, β_0 is the base value of attractiveness parameter at distance zero ($r=0$). For faster calculation, the exponential function in (5) if required, can be approximated as

$$\beta = \frac{\beta_0}{1+\gamma r^2} \quad (6)$$

Equation (5), defines a characteristic distance $\Gamma = 1 / \sqrt{\gamma}$ over which attractiveness varies considerably from β_0 to $\beta_0 e^{-1}$. If the attractive function $\beta(r)$, is implemented as a monotonically decreasing function, then $\beta(r)$, is defined as

$$\beta(r) = \beta_0 e^{-\gamma r^m}; m \geq 1 \quad (7)$$

The characteristics length $\Gamma = \gamma^{-1/m}$ will become 1 as $m \rightarrow \infty$ for a fixed value γ , conversely in the optimization problem for a length Γ , the parameter γ will be used as, $\gamma = 1 / \Gamma^m$. Let i and j be any two fireflies; then the distance between them at z_i and z_j is the Cartesian distance, define as

$$r_{ij} = |z_i - z_j| = \sqrt{\sum_{k=1}^d (z_{i,k} - z_{j,k})^2} \quad (8)$$

where $z_{i,k}$ is the k^{th} component of coordinate z_i of the i^{th} firefly.

The attraction of i^{th} firefly towards more brighter j^{th} firefly, is determined as-

$$z_i = z_i + \beta_0 e^{-\gamma r_{ij}^2} (z_j - z_i) + \alpha \left(\text{rand} - \frac{1}{2} \right) \quad (9)$$

where, the second term is due to attraction assumption and the third term is randomization with parameter α , 'rand' is a uniformly distributed random number in $[0, 1]$.

The proposed Firefly Algorithm implemented for optimal feature selection of masses in digital mammograms is as shown in Fig. 2.

First, an initial population of fireflies is created. The initial population includes possible solution. The number of fireflies in the initial population depends on the parameter, nPop. Each solution vector in the initial population is encoded using binary encoding technique. A variable that represents the possible solution is a 25-bit binary vector representing 25 features extracted from segmented masses. A bit value '1' in the solution vector indicates that the corresponding feature is selected while bit value '0' indicates that the feature is not

Algorithm : Feature selection using FA

Input : Twenty five features extracted from each of the 651 mammograms

1. Initialize FA parameters
 $MaxIT$ // maximum iterations
 $nPop$ // number of Fireflies
 $gama$ // light absorption coefficient
 $beta_o$ // attraction coefficient base value
 $alpha$ // mutation coefficient
2. Define objective function $f(z)$ // evaluate cost
3. Load feature data set // extracted features
4. Create initial population of fireflies z_i ($i=1,2,3,\dots,nPop$)
5. Evaluate initial population, $f(z)$ = accuracy of ANN/SVM
6. While (iteration<MaxIT)
7. for $i = 1$ to $nPop$ // firefly z_i
8. for $j = 1$ to $nPop$ // firefly z_j
9. if $f(z_j) > f(z_i)$ then
10. update r_{ij} //distance between firefly
11. update $beta$ //attractiveness coefficient
12. evaluate new solution using $f(z)$
13. if $f(\text{old_solution}) \leq f(\text{new_solution})$
14. then
15. Best_solution=new_solution
16. end if
17. end if
18. end of loop j
19. end of loop i
20. Rank the fireflies and find current best
21. end while loop
22. Obtain the final result as optimal features

Fig.2. Firefly Algorithm

selected. The fitness of each possible solution in the initial population is evaluated using objective function. The classification accuracy of ANN and SVM acts as an objective function (fitness). It is used to find the fitness of each selected feature. It is defined as –

$$\text{Objective function } f(z) = \text{mean}(\text{accuracy}) \quad (10)$$

Then, a new firefly is generated and evaluated. The best solution for the current iteration is obtained by comparing new and old solutions. Finally, the optimal feature subset obtained for the feature set is used for reduction.

The optimal feature subset selected by FA with ANN (FA-ANN), and FA with SVM (FA-SVM) are as shown in Table 2 and Table 3 respectively.

C. Classification

The proposed methodology uses Artificial Neural Network (ANN) and Support Vector Machine (SVM) for classification of masses in Digital mammograms into benign and malignant.

Table 2. FA-ANN based feature selection

Max. Iteration	Pop. Size	Feature Set	No. of Features Selected	Selected Features	Selection Accuracy (%)
10	15	AF1	11	2, 4, 6, 8, 11, 13, 17, 20, 21, 23, 24	95.23
15	15	AF2	12	1, 2, 5, 6, 8, 13, 16, 20, 21, 23, 24, 25	93.40
20	15	AF3	14	1, 5, 6, 8, 11, 13, 14, 16, 18, 19, 20, 22, 24, 25	93.70
30	15	AF4	14	1, 2, 5, 6, 8, 9, 10, 12, 14, 19, 21, 22, 23, 25	90.20
40	15	AF5	13	1, 3, 6, 11, 12, 14, 15, 17, 18, 19, 22, 23, 25	92.60

Table 3. FA-SVM based feature selection

Max. Iteration	Pop. Size	Feature Set	No. of Features Selected	Selected Features	Selection Accuracy (%)
10	15	SF1	13	1, 4, 6, 7, 8, 9, 10, 16, 17, 19, 20, 21, 23	90.90
15	15	SF2	12	1, 6, 7, 8, 15, 17, 18, 19, 20, 21, 24, 25	90.10
20	15	SF3	14	1, 2, 6, 8, 9, 10, 11, 14, 15, 18, 20, 21, 23, 25	92.47
30	15	SF4	9	1, 2, 4, 6, 8, 18, 19, 20, 22	89.10
40	15	SF5	13	1, 3, 6, 7, 8, 9, 12, 16, 17, 19, 20, 23, 24	89.25

1. Artificial Neural Network (ANN)

Artificial neural network is a simplified imitation of the central nervous system and thus, inspired by the types of computation performed by the human brain [24-26]. ANN is a massively parallel-distributed system made up of highly interconnected processing elements called nodes or neurons, working in unison to solve a specific problem. The major advantage of ANN is the ability to learn and adopt. In ANN learning involves adjustments of weights between neurons at the hidden and output layers. The behavior of an Artificial Neural Network depends on weights and input-output transfer function specified for the units. ANN generally uses non-linear transfer functions to generate the desired output [27-28].

The proposed method uses multilayer perceptron (MLP) for the classification of masses. It consists of three layers: input layer, hidden layer and output layer [29]. In this net,

information flows only in one direction (forward), hence MLP is known as feed forward neural network. The input layer simply passes the input to the next layer without any computation. Computation is performed at the hidden and output layers. Output layer ascertains the problem class (benign and malignant). Multilayer perceptron is trained using supervised training method. It requires training data comprising of a set of inputs with the associated outputs (targets). 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 weight adjustment for minimizing the overall error. Training of the network is repeated with the input data and weights adjusted until the desired input-output mapping is achieved [30]. In ANN, mean square error (MSE) determines the amount of weight change. Once the network is trained, validity of the model is evaluated by testing the model with unseen data.

2. Support Vector Machine (SVM)

Support vector machine is a supervised learning algorithm, developed by Vapnik, for solving classification problems [31]. For a two-class problem, the basic objective of SVM is to find an optimal hyperplane that separates two classes. A hyperplane is an optimal hyperplane if it separates the data with maximal margin. Data points close to maximal margin hyperplane are called support vectors, and the distance between data point and hyperplane is called margin of the SVM classifier. Let us consider a training set T with set of patterns and set of class labels, defined as –

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (11)$$

where, x_i is a pattern and $y_i \in \pm 1$ is a class label.

Let S be a dot product space, then a hyperplane in S , is defined as-

$$\{x \in S | w \cdot x + b = 0\}, w \in S, b \in R \quad (12)$$

where, w is a weight normal to the line, b is a bias and $w \cdot x = \sum_{i=1}^m w_i x_i$

The proposed method uses 'linear' Kernel based SVM with 10-fold cross validation. Therefore, SVM is a hyperplane T ($w \cdot x + b = 0$) with maximum margin between two hyper planes T_1 and T_2 , defined as

$$w \cdot x_i + b \geq +1 \text{ when } y_i = +1 \quad (13)$$

$$w \cdot x_i + b \leq -1 \text{ when } y_i = -1 \quad (14)$$

SVM classifies, data using Equation (15), defined as

$$\text{class}(x_i) = \begin{cases} +1 & \text{if } x_i \cdot w + b > 0 \\ -1 & \text{if } x_i \cdot w + b < 0 \end{cases} \quad (15)$$

IV. RESULTS AND DISCUSSION

The proposed Firefly algorithm based methodology was tested on over 651 digital mammograms obtained from Digital Database for Screening Mammography (DDSM) [32-33]. Out of 651 cases, 314 were benign and 337 were malignant. The process of feature selection by Firefly algorithm with ANN and SVM classifiers is discussed in Section III. FA uses the following parameters, selected by trial and error method.

Maximum iterations (<i>MaxIT</i>)	: 10/15/20/30/40/50
Number of Fireflies (<i>nPop</i>)	: 15
Light absorption coefficient (<i>gamma</i>)	: 0.5
Attraction coefficient (<i>beta₀</i>)	: 1
Mutation coefficient (<i>alpha</i>)	: 0.05

FA selects the subset of features based on fitness of the selected feature. Classification accuracy of ANN and SVM acts as the fitness value as defined in Equation (10). The subset of features selected by FA-ANN is as shown in Table 2. From Table 2, we observe that, FA-ANN selected five subsets AF1-AF5 for a fixed population size (Pop size=15) and varying number of iterations (Max. Iteration=10, 15, 20, 30, 40). The feature subset AF1 having 11 features appear to be the best, with a selection accuracy of 95.23%. The subset of features selected by FA-SVM is as shown in Table 3. From Table 3, we observe that FA-SVM selected five subsets of features, SF1-SF5. Subset SF3 appears to be the best feature subset having 14 features with a selection accuracy of 92.47%.

Feature subset AF1 acts as an input for the training and testing of artificial neural network. An ANN classifier is trained with the following parameters:

Hidden layer size	: 12
Training function	: 'trainlm'
No. of epoch	: 1000
Performance function	: 'mse'
Minimum gradient (min_grad)	: 1.0000e-07
Maximum failures (max_fail)	: 6
Momentum (mu)	: 1.0000e-03

The proposed ANN (Multilayer Perceptron) consists of eleven nodes at input layer, twelve nodes at hidden layer and two nodes at output layer. The input data of 651 samples was portioned as 70% for training, 15% for testing and 15% for validation. The network trained for 1000 epoch using Levenberg-Marquardt ('trainlm') method. The best validation performance of the network obtained is 0.15707 at an epoch 23, as shown in Fig. 3.

The gradient and momentum parameters achieved by the network during training areas shown in Fig. 4. From Fig. 4, we observe that gradient and momentum parameter values obtained at epoch 29 is 0.019648 and 0.01 respectively.

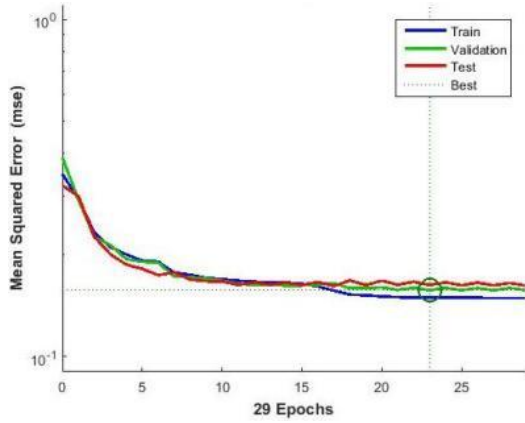


Fig.3. Performance of ANN

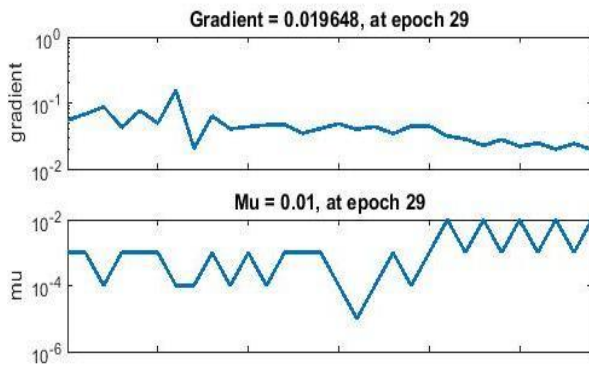


Fig.4. ANN training parameter values

The feature subset SF3 acts as input for the training and testing of SVM classifier. The classifier SVM was trained by 'linear' kernel function with 10-fold cross validation. The 10-fold cross validation technique was used to ensure correctness of results produce by the classifier. In 10-fold cross validation, input samples were randomly divided into 10-subsamples. Out of 10-subsamples, 9-subsamples were used to train the classifier and one subsample was used to validate the classifier. Result of the classifier is generated by averaging the results of all 10-folds validations. The parameters of SVM classification model, sigma and momentum per feature are as shown in Fig. 5. Other parameter values with Sequential Minimal Optimization (SMO) solver are, bias 0.2486 and kernel scale 0.2486.

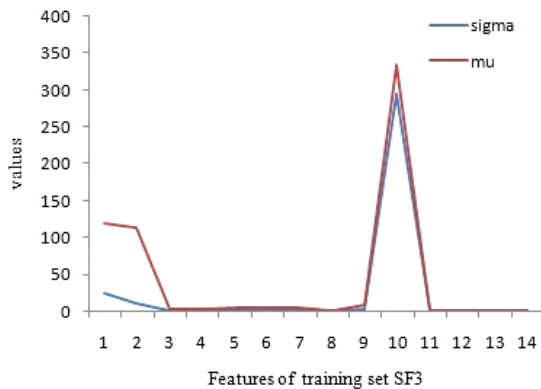


Fig.5. SVM model parameters

The performance of the classifiers, evaluated using statistical parameter is as listed in Equations (16)-(21).

Sensitivity (TPR): It defines the amount of malignant (positive) cases, correctly classified as true positive (TP), out of all positive cases.

$$TPR = \frac{TP}{TP+FN} \quad (16)$$

Specificity (TNR): It defines the amount of benign (negative) cases correctly classified as true negative (TN), out of all negative cases.

$$TNR = \frac{TN}{TN+FP} \quad (17)$$

Accuracy (ACC): It defines the total amount of TP and TN cases correctly classified as TP and TN, out of all the cases.

$$ACC = \frac{TP+TN}{TP+FN+TN+FP} \quad (18)$$

Type-I error: Type-I error is also known as false alarms. It is define as false positive rate (FPR).

$$FPR = \frac{FP}{FP+TN} \quad (19)$$

Mean square error (MSE): The weights and bias values of the network are, updated according to mean square error. It is the difference between target and the output generated by the network.

$$MSE = \frac{1}{n} \sum_1^n (O_i - T_i)^2 \quad (20)$$

Root mean square error (RMSE): It is defined as-

$$RMSE = \sqrt{MSE} \quad (21)$$

Area under ROC Curve (AUC): The area under receiver operating characteristics (ROC) curve is a plot between sensitivity (TPR) on Y-axis and 1-Specificity on the X-axis [34].

The summary of classifier performance is as shown in Table 4. From Table 4, we observe that artificial neural network (ANN) is superior to support vector machine (SVM). ANN achieves an accuracy of 95.23% with a sensitivity of 96.43% and specificity of 93.94%. On the other hand, SVM achieves an accuracy of 92.47% with 96.14% sensitivity and 88.53% specificity. The misclassification rate of ANN and SVM are 4.77% and 7.53% respectively. Similarly, false alarm or type-I error (FPR) for ANN and SVM is 3.14% and 11.46% respectively.

Table 4. Classifier performance

Methods	TP	FN	TN	FP	TPR (%)	TNR (%)	ACC (%)
ANN	325	12	295	19	96.43	93.94	95.23
SVM	324	13	278	36	96.14	88.53	92.47

TP: True Positive, FN: False Positive, TN: True Negative, FP: False Positive, TPR: True Positive Rate, TNR: True Negative Rate, ACC: Accuracy

Other parameters, mean square error (MSE), root mean square error (RMSE) and correlation coefficient (R), measure the predictive power of a classification model. An error is defined as the difference between target and the output generated by the classifier. The error of ANN and SVM per training sample is as shown in Fig. 6 and Fig. 7. The mean square error of ANN and SVM is 0.08013 and 0.0860 respectively. Smaller the value of MSE better is the model. Root mean square is the square root of the difference between target and output of classifier. RMSE Value of ANN and SVM is 0.2141 and 0.2933 respectively. Similarly, the correlation coefficient (R) between target and output for training, testing and validation of ANN is 0.85082. The classification model is 100% linear, if $R=1$. It is concluded that ANN classifier is superior to SVM with reference to MSE and RMSE parameters.

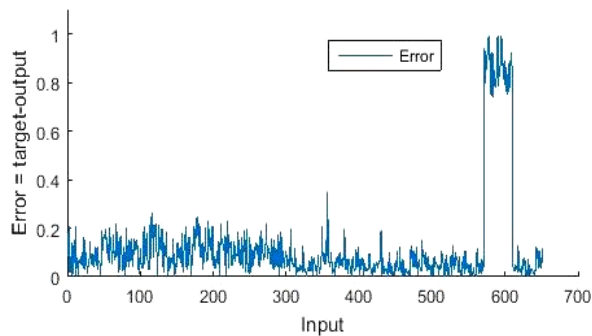


Fig.6. ANN error per training sample

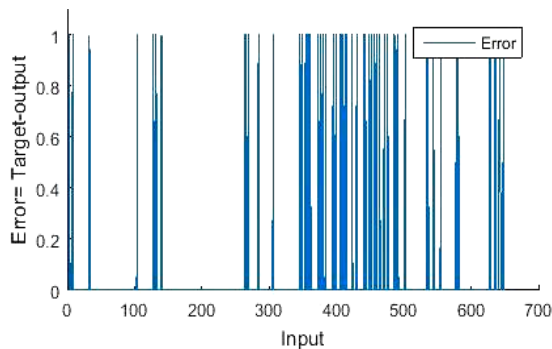


Fig.7. SVM error per training sample

Another useful parameter used to determine the performance of the classifiers is the area under receivers

operating characteristics (ROC) curve. It is a graph between true positive rate (TPR) and false positive (FPR) rate. The area under ROC curve lies between 0 and 1. If area under curve is equal to one, classifier is said to be 100% correct [34]. The receiver operating characteristics curves for ANN and SVM are as shown in Fig. 8 and the area under ROC curve with 95% Confidence interval for ANN and SVM is presented in Table 5.

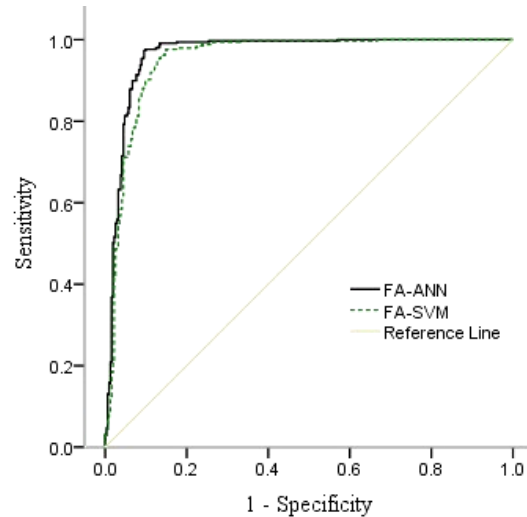


Fig.8. Comparison of ROC curves of ANN and SVM classifiers

Table 5. Area under ROC curve for ANN and SVM

Method	Area	Std. Error	Asymptotic sig.	Asymptotic 95% Confidence Interval	
				LB	UB
FA-ANN	.965	.008	.000	.950	.980
FA-SVM	.951	.009	.000	.933	.969

From Table 5, we observe that the area under curve for ANN is $A_z=0.965 \pm 0.008$ and for SVM it is $A_z=0.951 \pm 0.009$. From Fig. 8, we can see that ROC curve for ANN is close to one.

Performance of the classifier depends on the feature selection process as well as the number of cases used for training of classifiers. Use of very few and unbalanced cases (benign and malignant) for training of classifiers would affect the performance of the classifier as well as the feature selection process [35-36]. The proposed firefly based feature selection and classification method used a large and balanced data set, comprising of 651 cases.

With the above discussion, we conclude that FA-ANN method is superior to FA-SVM, in terms of statistical and other parameters used for evaluating the performance of the classifiers. The method, FA with ANN, selects the most optimal features with an accuracy of 95.23%. The proposed method can help to improve breast cancer diagnosis.

V. CONCLUSION

In this paper, we studied methods for feature selection and classification of masses in digital mammograms. Feature selection is carried out by Firefly algorithm, with artificial neural network and support vector machine. Feature selection by firefly algorithm depends on fitness value. Classification accuracy of ANN and SVM acts as fitness value. Features selected by FA-ANN and FA-SVM were used to train and test two classifiers, artificial neural network and support vector machine. Performance study of the proposed methods showed that FA-ANN based feature selection and classification technique is superior to FA-SVM. Result achieved with the proposed method proves that Firefly algorithm can be effectively applied for feature selection and classification of masses. This method can help to improve breast cancer treatment and would minimize mortality rate.

REFERENCES

- [1] (NCI), N. C. I. Cancer stat fact sheets: Cancer of the breast. Available at: <http://www.seer.cancer.gov/statfacts/html/breast.html>, May 2009.
- [2] B. Acha, C. Serrano, R.M. Rangayyan and J.L. Desautels, "Detection of microcalcifications in mammograms," *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*. SPIE, Bellingham, 2006.
- [3] C. Blum and Li. Xiaodong, "Swarm intelligence in optimization," In *Swarm Intelligence*, pp. 43-85. Springer Berlin Heidelberg, 2008.
- [4] D. Karaboga and B. Bahriye, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," *Journal of global optimization*, Vol.39, pp. 459-471, 2007.
- [5] M. Dorigo and T. Stutzle, "The ant colony optimization metaheuristic: Algorithms, applications, and advances," *International series in operations research and management science*, pp. 251-286, 2003.
- [6] J. Kennedy and R.C. Eberhart, "The particle swarm: social adaptation in information-processing systems," In *New ideas in optimization*, pp. 379-388. McGraw-Hill Ltd., UK, 1999.
- [7] X. S. Yang, "Nature-inspired metaheuristic algorithms," Luniver press, UK, pp. 242-246, 2008.
- [8] X. S. Yang, "Firefly algorithms for multimodal optimization," In *International symposium on stochastic algorithms*, pp. 169-178, Springer, Berlin, Heidelberg, 2009.
- [9] X. S. Yang and S. Deb, "Cuckoo search via Lévy flights," In *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*, pp. 210-214, IEEE, 2009.
- [10] X. S. Yang, "A new metaheuristic bat-inspired algorithm," *Nature inspired cooperative strategies for optimization*, pp. 65-74, 2010.
- [11] A. H. Gandomi and A. H. Alavi, "Krill herd: a new bio-inspired optimization algorithm," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, pp. 4831-4845, 2012.
- [12] X. S. Yang and X. He, "Firefly algorithm: recent advances and applications," *International Journal of Swarm Intelligence*, vol. 1, pp. 36-50, 2013.
- [13] E. Saraç and S. A. Özel, "Web page classification using firefly optimization," In *Innovations in Intelligent Systems and Applications (INISTA), 2013 IEEE International Symposium on*, pp. 1-5, IEEE, 2013.
- [14] J. Senthilnath, S. N. Omkar and V. Mani, "Clustering using firefly algorithm: performance study," *Swarm and Evolutionary Computation*, vol. 1, pp.164-171, 2011.
- [15] H. Banati and M. Bajaj, "Firefly based feature selection approach," *International Journal of Computer Science*, vol. 8, pp. 473-480, 2011.
- [16] C. L. Blake, "UCI repository of machine learning databases," <http://www.ics.uci.edu/~mllearn/MLRepository.html>, 1998.
- [17] S. K. Pal, C. S. Rai and A. P. Singh, "Comparative study of firefly algorithm and particle swarm optimization for noisy non-linear optimization problems," *International Journal of intelligent systems and applications*, vol. 4, pp. 50, 2012.
- [18] G. P. Singh and A. Singh, "Comparative study of Krill Herd, firefly and cuckoo search algorithms for unimodal and multimodal optimization," *International Journal of Intelligent Systems and Applications*, vol. 6, pp. 35, 2014.
- [19] K. G. Dhal, I. Quraishi and S. Das, "A Chaotic Lmeta," *International Journal of Image, Graphics and Signal Processing*, vol. 7, pp. 69, 2015.
- [20] S. Thawkar and R. Ingolikar, "Automatic Detection and Classification of Masses in Digital Mammograms," *International Journal of Intelligent Engineering and Systems*, vol. 10, pp. 65-74, 2017.
- [21] S. Thawkar and R. Ingolikar, "Efficient approach for the classification of masses in digital mammograms," *International journal of innovative computing information and control*, vol.13, pp. 967-978, 2017.
- [22] M. Sameti, R. K. Ward, J. Morgan-Parkes and B. Palcic, "A method for detection of malignant masses in digitized mammograms using a fuzzy segmentation algorithm," In *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, vol. 2, pp. 513-516, IEEE, 1997.
- [23] H. Li, Y. Wang, K. R. Liu, S. C. Lo and M. T. Freedman, "Computerized radiographic mass detection. II. Decision support by featured database visualization and modular neural networks," *IEEE transactions on medical imaging*, vol. 20, pp. 302-313, 2001.
- [24] K. Bovis, S. Singh, J. Fieldsend and C. Pinder, "Identification of masses in digital mammograms with MLP and RBF nets," In *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, vol. 1, pp. 342-347, IEEE, 2000.
- [25] S. Baeg, and N. Kehtarnavaz, "Texture based classification of mass abnormalities in mammograms," In *Computer-Based Medical Systems, 2000. CBMS 2000. Proceedings. 13th IEEE Symposium on*, pp. 163-168, IEEE, 2000.
- [26] R. P. Velthuizen and J. I. Gaviria, "Computerized mammographic lesion description," In *[Engineering in Medicine and Biology, 1999. 21st Annual Conference and the 1999 Annual Fall Meeting of the Biomedical Engineering Society] BMES/EMBS Conference, 1999. Proceedings of the First Joint*, vol. 2, pp. 1034-vol, IEEE, 1999.
- [27] D. B. Fogel, E. C. Wasson III, E. M. Boughton and V. W. Porto, "Evolving artificial neural networks for screening features from mammograms," *Artificial Intelligence in Medicine*, vol. 14, pp. 317-326, 1998.

- [28] C. E. Floyd, J. Y. Lo, A. J. Yun, D. C. Sullivan and P. J. Kornguth, "Prediction of breast cancer malignancy using an artificial neural network," *Cancer*, vol. 74, pp. 2944-2948, 1994.
- [29] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai and H. N. Du, "Approaches for automated detection and classification of masses in mammograms," *Pattern recognition*, vol. 39, pp. 646-668, 2006.
- [30] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric environment*, vol. 32, pp. 2627-2636, 1998.
- [31] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273-297, 1995.
- [32] M. Heath, K. Bowyer, D. Kopans, R. Moore and W. P. Kegelmeyer, "The digital database for screening mammography," In *Proceedings of the 5th international workshop on digital mammography*, pp. 212-218, Medical Physics Publishing, 2000.
- [33] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer Jr, R. Moore, K. Chang and S. Munishkumaran, "Current status of the digital database for screening mammography," In *Digital mammography*, pp. 457-460, Springer Netherlands, 1998.
- [34] J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol. 240, pp. 1285-1293, 1988.
- [35] M. Kubat, R. C. Holte and S. Matwin, "Machine learning for the detection of oil spills in satellite radar images," *Machine learning*, vol. 30, pp. 195-215, 1998.
- [36] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: one-sided selection," In *ICML*, vol. 97, pp. 179-186, 1997.

Authors' Profiles



Shankar Thawkar received his master degree in Computer Science and Engineering from Dr. B. R. Ambedkar University, Agra, India. He is presently working as an Asst. Prof. in the department of information technology, Hindustan College of science and technology, Mathura, India. His area of research includes medical image processing, artificial intelligence, machine learning and pattern recognition. He has published more than eight research papers in National/International Journals and Conferences.



Ranjana Ingolikar received her Ph.D. degree in Computer Science from RTM university, Nagpur, India. She is presently working as a HOD, Department of Computer Science S. F. S. College, Nagpur, India. Her area of research includes artificial intelligence, machine learning, security, data mining and object oriented system. She has published many Research papers in National/ International Journals and Conferences.

How to cite this paper: Shankar Thawkar, Ranjana Ingolikar, "Classification of Masses in Digital Mammograms Using Firefly based Optimization", *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, Vol.10, No.2, pp. 25-33, 2018. DOI: 10.5815/ijigsp.2018.02.03