

<http://www.pjbs.org>

PJBS

ISSN 1028-8880

**Pakistan
Journal of Biological Sciences**

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Ant-cuckoo Colony Optimization for Feature Selection in Digital Mammogram

¹J.B. Jona and ²N. Nagaveni

¹Department of CTA,

²Department of Mathematics, Coimbatore Institute of Technology, India

Abstract: Digital mammogram is the only effective screening method to detect the breast cancer. Gray Level Co-occurrence Matrix (GLCM) textural features are extracted from the mammogram. All the features are not essential to detect the mammogram. Therefore identifying the relevant feature is the aim of this work. Feature selection improves the classification rate and accuracy of any classifier. In this study, a new hybrid metaheuristic named Ant-Cuckoo Colony Optimization a hybrid of Ant Colony Optimization (ACO) and Cuckoo Search (CS) is proposed for feature selection in Digital Mammogram. ACO is a good metaheuristic optimization technique but the drawback of this algorithm is that the ant will walk through the path where the pheromone density is high which makes the whole process slow hence CS is employed to carry out the local search of ACO. Support Vector Machine (SVM) classifier with Radial Basis Kernel Function (RBF) is done along with the ACO to classify the normal mammogram from the abnormal mammogram. Experiments are conducted in miniMIAS database. The performance of the new hybrid algorithm is compared with the ACO and PSO algorithm. The results show that the hybrid Ant-Cuckoo Colony Optimization algorithm is more accurate than the other techniques.

Key words: Cuckoo search, ant colony optimization, ROC curve, support vector machine

INTRODUCTION

Breast cancer is the second leading causes of mortality in women. So far mammography is the only effective screening method for detection of breast cancer in early stage. There exist two errors like false negative and false positive which may arise in interpretation of mammography by the radiologist. To overcome such errors in the interpretation of mammography the researchers developed Computer Aided Diagnosis (CADx) which evaluate or assess mammographic abnormality by automating the segmentation, detection, feature extraction and classification processes.

During the past two decades the focus of researchers falls on the nature inspired metaheuristic algorithms. They concentrated more on the Nature-Inspired Computation (NIC). The NIC are computational algorithms that are mimicked from the natural phenomena and biological models of animals and birds to solve a problem. Many nature inspired algorithms had evolved in the past few decades. The most well-known NIC are the artificial neural networks, Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Artificial Immune System (AIS).

ACO is a population based optimization technique which was first introduced by Marco Dorigo, as Ant System (Dorigo *et al.*, 1996; Dorigo and Gambardella, 1997a). In 1999 it was redefined as Ant Colony

Optimization and used first to solve the travelling salesman problem (Dorigo and Stutzle, 2004; Dorigo and Gambardella, 1997b). It is inspired by the behavior of ants in finding shortest paths from the colony to its food. Al-Ani used ACO for feature selection (Al-Ani, 2005). Cuckoo Search (CS) is an optimization technique developed by Xin-She Yang and Suash Deb (Yang and Deb, 2009, 2010; Tuba *et al.*, 2011) which was inspired from the egg laying and breeding behavior of the cuckoo bird. Cuckoo search is simple and takes few parameters compared to other metaheuristic algorithms. They are proved that "Cuckoo search outperforms PSO". The major disadvantage of ACO is that the local search it performs is not much faster. So Cuckoo search is proposed to carry out the local search of ACO. In this paper a hybrid of Ant Colony Optimization (ACO) and Cuckoo Search are proposed.

AN OVERVIEW OF CUCKOO SEARCH ALGORITHM

Cuckoo Search (CS) is population based metaheuristic optimization technique. The basic behavior of cuckoo is that it lays its eggs in some other bird's (host bird's) nest. If the cuckoo's egg is similar to the host bird's egg then the egg is not discovered by the host bird. If the egg is recognized by host bird then the egg will be

destroyed by the host bird. So the cuckoo mimics the color and pattern of their own eggs to match that of their host. If the cuckoo's egg hatches first then the cuckoo chick destroy the eggs of the host. By imitating the egg laying behavior of the cuckoo an optimization algorithm has been developed.

Cuckoo optimization algorithm is based on three simple rules stated:

- Cuckoo lays only one egg at a time and it dumps its egg in a randomly chosen nest
- Only the best nests with high quality eggs will be passed into the next generation
- The number of available host nests is fixed. Egg laid by a cuckoo bird is discovered by the host bird with a probability P_d . In this case, the host bird has two options. It can either throw the egg away, or it may abandon the nest and build a brand new nest at nearby location

CUCKOO SEARCH FOR FEATURE SELECTION

In cuckoo search optimization algorithm each egg in a nest represents a solution and a cuckoo egg represents a new solution. The aim of the CS is to use the new and potentially better solutions to be replaced by not-so-good solution in the nests. In the simplest form, each nest has one egg. The host nest is represented by N and the number of cuckoos is referred as S . Each cuckoo's egg represents the selected feature. For each host nest, the quality of eggs is either 1 or 0, which represent whether the selected feature will carry on to the generation or not. i.e. whether that particular feature is selected or not. Probability that the egg laid by a cuckoo is discovered by the host bird is P_d . A minimum value of P_d leads to discarding the feature and high value of P_d takes that particular feature to the next generation. The fitness of a nest is obtained by evaluation of fitness function f_i .

To start the optimization algorithm, a candidate matrix of size $S \times N$ is generated. Initialization of the nests and the random initial solution is performed. Evaluate the fitness of the initial solution and get the current best solution. The current best solution performed using Eq. 1:

$$\sigma_u = \left[\frac{\Gamma(1+\beta) \sin\left(\frac{\alpha\pi}{2}\right)}{\Gamma\left[\frac{(1+\beta)}{2}\right] \beta_2 \left(\frac{\beta-1}{2}\right)} \right]^{\frac{1}{\beta}}, \sigma_v = 1 \quad (1)$$

where, β denotes Levy distribution parameter, $\beta=1.5$ and Γ denotes gamma function. Generate new solution but keep the current best nest. A new solution $x^{(t+1)}$ for cuckoo i is generated using a Lévy flight:

$$x_i^{(t+1)} = x_i^{(t)} + \alpha r^s \quad (2)$$

where, $\alpha > 0$ is the step size, in most cases $\alpha = 1$ and \wedge represents entry-wise multiplications. The algorithm perform random walk using the Levy flight of Mantegna's algorithm with step size. Step size is calculated using Eq. 3. From the normal distribution:

$$\text{Stepsize}^{(t+1)} = 0.001 \frac{u^{(t+1)}}{|v^{(t+1)}|^{\frac{1}{\beta}}} (v - x_{\text{best}}) \quad (3)$$

Evaluate the new solution to obtain the current best nest. Destroy the nest which has the probability less than P_d and retain remaining nests to the next generation. Perform the new nest generation by retaining the current best nest of the previous generation, evaluate the fitness of each nest, find current best nest and eliminate the nest which has a value less than P_d . All these steps are repeated until it meets the maximum number of generations. The pseudo code of the cuckoo search is given below.

Algorithm 1: CS algorithm

-
1. Initial a population of n host nests x_i ($i = 1, 2, \dots, n$);
 - Initialize the random solution.
 - Choose the current best nest by evaluating the fitness
 2. while ($t < \text{MaxGeneration}$)
 - Get a cuckoo (say i) randomly by Levy flights
 - Evaluate its quality/fitness f_i
 - Choose a nest among n randomly
 - if ($f_i > f_j$),
 - 2.4.1 Replace j by the new solution
 - Abandon a fraction (P_d) of worse nests
 - 2.5.1 build new ones at new locations via Levy flights
 - Keep the best solutions
 - Rank the solutions and find the current best
 - Repeat step 2
-

The advantage of the Cuckoo search is that it takes only one parameter in addition to the population size. Cuckoo search is so simple and more efficient than other metaheuristic algorithms. The randomization is more efficient as the step length is heavy-tailed and any large step is possible.

METHODOLOGY

The mammographic images from MiniMIAS database is used in this research. Seventy eight Gray Level Co-occurrence Matrix (GLCM) textural features are extracted. The extracted feature set is reduced by the proposed hybrid Ant-Cuckoo Colony optimization technique. SVM classifier is used along with ACO to classify the normal mammogram from an abnormal mammogram. The performances of all the proposed techniques are compared by Receiver Operating Characteristic curve (ROC).

Feature extraction: Feature extraction is the process of reducing the original mammogram image into a set of

features, by measuring certain properties or features that distinguish one input pattern from another pattern. GLCM features are extracted from the mammogram. GLCM features are calculated based on the haralick's texture feature. The haralick features are namely energy, correlation, inertia, entropy, inverse difference moment, sum average, sum variance, sum entropy, difference average, difference variance, difference entropy, information measure of correlation 1 and information measure of correlation 2 are extracted at four directions (0°, 45°, 90°, 135°). The mean and variance of each of the thirteen haralick feature at four directions are extracted making a total of 78 features (Haralick *et al.*, 1973).

Feature selection using ant-cuckoo colony optimization:

ACO is inspired by the foraging behavior of real ants. While walking from food sources to the nest and vice versa, ants deposit a chemical substance called pheromone on the ground. When they decide about a direction to go, they choose probabilistically paths marked by strong pheromone concentrations. This behavior is the basis for a cooperative interaction which leads to the emergence of shortest paths between food sources and their nest. In ACO algorithms, the artificial ants incrementally construct a solution by adding appropriately defined solution components to the current partial solution. Each of the construction steps is a probabilistic decision based on local information, which is represented by the pheromone information.

The pseudo code of the ACO algorithm is shown in Algorithm 2. The algorithm is initialized with the No. of features, No. of iteration, No. of ants, No. of selected features, trail intensity (T_i), evaporation rate, etc., In the first iteration, each ant will randomly choose a feature subset of m features. Only the best k subsets, ($k < na$), will be used to update the pheromone trail and influence the feature subsets of the next iteration. In the second and following iterations, each ant will start with $m-p$ features that are randomly chosen from the previously selected k -best subsets, where p is an integer that ranges between 1 and $m-1$. This process of feature selection is a sequential process, which may delay the entire system of feature selection. So the Cuckoo search of Algorithm 1 selects the optimal feature. The feature selected by the cuckoo search is now carried to the next generation. In this way, the features that constitute the best k subsets will have more chance to be present in the subsets of the next iteration. However, it will still be possible for each ant to consider other features as well. For a given ant j , those features are the ones that achieve the best compromise between previous knowledge, i.e., pheromone trails and the current best of the cuckoo search.

The problem of feature selection can be stated as follows: given the feature set, F , of n features, find the feature subset S , which consists of m features where $m < n$ and $S \subset F$, such that the classification accuracy is maximized. The feature selection problem representation exploited by the artificial ants includes the following:

- n features that constitute the original set, $F = \{f_1, \dots, f_n\}$
- na , the number of artificial ants to search through the feature space
- T_i , the intensity of pheromone trail associated with feature f_i
- For each ant j , a list that contains the selected feature subset, $S_j = \{s_1, \dots, s_m\}$

Algorithm 2: Ant-Cuckoo Colony Optimization algorithm for feature selection

1. Initialization:
 - Set $T_i = cc$ where cc is a constant
 - $\Delta T_i = 0$, where $i = 1, \dots, n$ and T_i is the amount of change of pheromone trail quantity for feature f_i .
 - Define the maximum number of iterations.
 - Define k , where the k -best subsets will influence the subsets of the next iteration.
 - Define m , the number of features to select
 - Define na , number of ants
 - Define p , where $m-p$ is the number of features each ant will start with in the second and following iterations.
2. First iteration,
 - For $j = 1$ to na ,
 - Randomly assign a subset of m features to S_j .
 - Go to step 4.
3. Select the remaining p features for each ant:
 - For $mm = m-p + 1$ to m ,
 - For $j = 1$ to na ,
 - Given subset S_j , choose feature f_i from the current best of cuckoo search of algorithm 1.
 - $S_j = S_j \cup \{f_i\}$.
 - Replace the duplicated subsets, if any, with randomly chosen subsets.
4. Evaluate the selected subset of each ant using a SVM classification algorithm:
 - For $j = 1$ to na ,
 - Estimate the Mean Square Error (MSE _{j}) of the classification results obtained by classifying the features of S_j .
 - Sort the subsets according to their MSE. Update the minimum MSE (if achieved by any ant) and store the corresponding subset of features.
5. Using the feature subsets of the best k ant:
 - For $j = 1$ to k ,

$$\Delta T_i = \begin{cases} \frac{\max(MSE_{e-1k}) - MSE_j}{e-1k} & \text{if } f_i \in S_j \\ \frac{\max(\max(MSE_{e-1k}) - MSE_j)}{n-k} & \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

$$T_i = \rho \cdot T_i + \Delta T_i \quad (5)$$

where ρ is a constant such that $(1-\rho)$ represents the evaporation of pheromone trails.

- For $j = 1$ to na ,
 - Randomly produce $m-p$ feature subset for ant j , to be used in the next iteration and store it in S_j .
6. If the number of iterations is less than the maximum number of iterations, go to step 3.

Support vector machine classifier: Support Vector Machine (SVM) is a classification techniques based on statistical learning theory (Bevilacqua *et al.*, 2000; El-Naqa *et al.*, 2002). SVM are Statistical Learning Theory (SLT) problems used for classification. The SVM algorithm constructs a separating hypersurface in the input space by transforming the input space into a high dimensional feature space through some nonlinear mapping chosen a priori (Kernel). It constructs the maximal margin hyperplane in the feature space and the support vectors that lies on this hyperplane. The discriminant function of SVM using the Kernel function is:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, y_i) + b \quad (6)$$

where, x_i is the support vector, y_i is the classification output (+1 for benign and -1 for malignant), α_i and b are quadratic programming coefficients. Radial Basis Function Kernel (RBF) is used and it is defined as:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (7)$$

where γ , is kernel parameter.

EXPERIMENTAL RESULTS AND DISCUSSION

Image database: In this research, mammograms from the Mammographic Image Analysis Society (MIAS), a Mini Mammographic Database is used. Each mammogram image has a spatial resolution of 1024×1024 pixels. This database is chosen since it contains various types of abnormalities such as calcification, well-defined, circumscribed masses, spiculated masses, ill-defined masses, architectural distortion, asymmetry and normal. Each of these abnormalities has been diagnosed and confirmed by a biopsy.

Experimental setup: The experiments implemented in MATLAB. These techniques are experimented on 100 mammogram images with various abnormalities, 50 abnormal images with microcalcification, spiculation, circumscribed and 50 normal mammogram images. Table 1 shows the parameters used in the algorithms. The parameters of the Ant-Cuckoo Colony Optimization algorithm are listed below.

Experimental results: The features selected by ACO, PSO and the proposed Ant-Cuckoo Colony Optimization are listed in Table 2. The 78 GLCM features are considered for the experiment and only the best 5 features are extracted.

Table 1: Parameters of Ant-Cuckoo colony optimization

Cuckoo parameters	Values	ACO parameters	Values
Step size α	1	Population size	100
Discovery rate P_d	0.25	Number of ant	78
Population size N	100	Number of Iterations	200
Max. generation	200	Evaporation rate ρ	0.75
		Heuristic value η	1
		α	1.0
		β	1.0

Table 2: Features selected by ACO, PSO and Ant-Cuckoo colony optimization

Techniques	Selected GLCM feature
ACO	Correlation, difference average, energy, inertia, sum variance
PSO	Correlation, energy, inverse difference, information measure of correlation 2, sum variance
Ant-Cuckoo colony optimization	Correlation, difference variance, information measure of correlation2, inertia, sum variance

With the selected features the testing is performed on the same set of 100 mammograms. The SVM classification results are pictorially depicted in Fig. 1.

Receiver Operating Characteristic (ROC) is a statistical tool used in medical decision making and is graphical plot of Sensitivity Vs 1-Specificity for a classifier. Sensitivity is the ratio of malignant samples which have a positive test result. Specificity is ratio of benign samples which have a negative test result. The True Positive Fraction (TPF) (or True Positive Ratio (TPR) or Sensitivity), False Positive Fraction (FPF) (or False Positive Ratio (FPR) or 1-Specificity) and accuracy are defined as:

$$TPF = \frac{TP}{TP + FN} \quad (8)$$

$$FPF = \frac{FP}{FP + TN} \quad (9)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (10)$$

Where:

TP = True positives is the No. of mammograms correctly classified as abnormal mammogram

FP = False positive is the No. of mammograms incorrectly classified as abnormal mammogram

TN = True negative is the No. of mammograms correctly classified as normal mammogram

FN = False negatives is the No. of mammograms incorrectly classified as normal mammogram

The confusion matrix is constructed from the obtained results is shown in Table 3. The accuracy of the hybrid technique is shows in Table 4. The results are compared with the traditional ACO and PSO. Figure 2 plots the excel graph to compare the performances of the

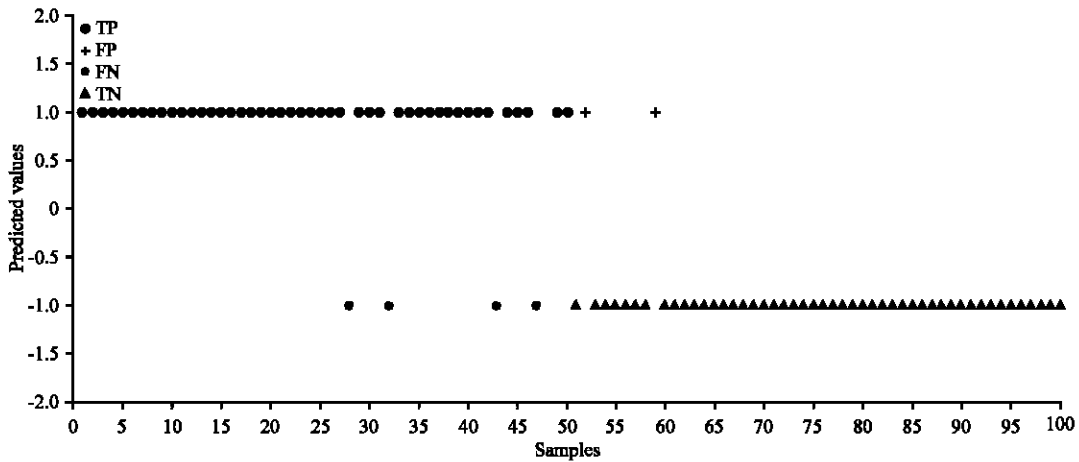


Fig. 1: Result of Ant-Cuckoo Colony Optimization-SVM

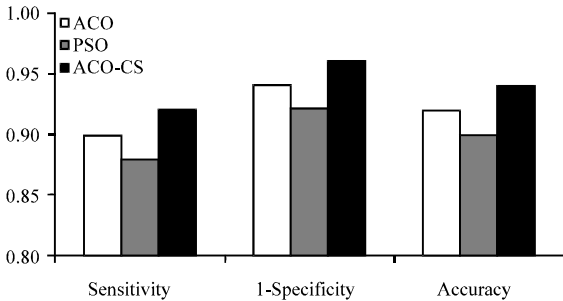


Fig. 2: Performance measure of the proposed techniques

Table 3: Confusion matrix

Technique	Actual	Predicted	
		Abnormal	Normal
ACO	Abnormal	(TP) 45	(FP) 3
	Normal	(FN) 5	(TN) 47
PSO	Abnormal	44	4
	Normal	6	46
ACCO	Abnormal	46	2
	Normal	4	48

Table 4: Performance of the proposed techniques

Classifier	TPF	TNF	Accuracy (%)
ACO	0.9	0.94	92
PSO	0.88	0.92	90
ACCO	0.92	0.96	94

algorithms. ROC curve plotted with TPR against FPR in Fig. 3 compares the accuracy percentage of all the three classifiers.

From Table 2, it is inferred that the features correlation and sum variance are selected by all the techniques. Energy, inertia and information measures of correlation 2 are the features selected by almost two of the techniques. The overall performance of the hybrid techniques is better than both PSO and ACO.

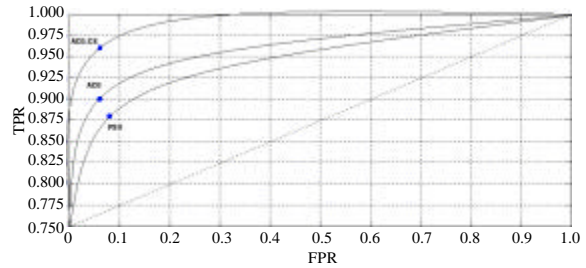


Fig. 3: ROC curves of ACO, PSO and Ant-Cuckoo colony optimization

Ant-Cuckoo Colony Optimization shows 2% better accuracy than ACO and 4% better accuracy than PSO.

CONCLUSION

This study presents the hybrid of Ant-Cuckoo colony optimization to select the best feature of digital mammogram. The local search of ACO, stimulated by the cuckoo search works well and gives an accuracy of 95%. The SVM classifier is the added feature of the proposed system. The SVM classifier plays dual role; it goes along with the Ant-Cuckoo Colony optimization and also used in testing the performance of total system. The hybrid algorithm shows promising accuracy and is better than the ACO and PSO algorithm.

REFERENCES

Al-Ani, A., 2005. Feature subset selection using ant colony optimization. *Int. J. Comput. Intell.*, 2: 53-58.

- Bevilacqua, A., D. Bollini, R. Brancaccio, R. Campanini and N. Lanconelli, 2000. Automatic detection of clustered microcalcifications in digital mammograms using an SVM classifier. Proceedings of the European Symposium on Artificial Neural Networks, April 26-28, 2000, DBLP., Bruges, Belgium, pp: 195-200.
- Dorigo, M., V. Mamezzo and A. Colomi, 1996. Ant system: Optimization by a colony of cooperating agents. IEEE Trans. Syst. Man Cybern. Part B: Cybern., 26: 29-41.
- Dorigo, M. and L.M. Gambardella, 1997a. Ant colonies for the traveling salesman problem. Bio Systems, 43: 73-81.
- Dorigo, M. and L.M. Gambardella, 1997b. Ant colony system: A cooperative learning approach to the traveling salesman problem. IEEE Trans. Evol. Comput., 1: 53-66.
- Dorigo, M. and T. Stutzle, 2004. Ant Colony Optimization. MIT Press, Cambridge, MA., USA.
- El-Naqa, I., Y. Yang, M.N. Wernick, Nikolas and P. Galatsanos *et al.*, 2002. A support vector machine approach for detection of microcalcifications. IEEE Trans. Med. Imaging, 21: 1552-1563.
- Haralick, R.M., K. Shanmugam and I. Dinstein, 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybernet., 3: 610-621.
- Tuba, M., M. Subotic and N. Stanarevic, 2011. Modified cuckoo search algorithm for unconstrained optimization problems. Proceedings of the 5th European Conference on European Computing, April 28-30, 2011, Paris, France, pp: 263-268.
- Yang, X.S. and S. Deb, 2009. Cuckoo search via Levy flights. Proceedings of the World Congress on Nature and Biologically Inspired Computing, December 9-11, 2009, Coimbatore, India, pp: 210-214.
- Yang, X.S. and S. Deb, 2010. Engineering optimization by cuckoo search. Int. J. Math Model. Num. Optim., 1: 330-343.