International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Implementation of Hybrid ACO-PSO-GA-DE Algorithm for Mammogram Classification

Anju Bala, Priti

Abstract: ***Breast Cancer is one of the fastest growing cancer that causes women to death in the world. The early detection of breast cancer improves the chances of its cure. The malignant tumor that is the sign of breast cancer can be detected by mammography. This paper develops a technique to classify the mammogram images as normal, benign or malignant. This paper applies HAPGD (Hybrid ACO (Ant Colony Optimization), PSO (Particle Swarm Optimization), GA (Genetic Algorithm), and DE (Differential Evolution)) classification algorithm to texture features extracted from the mammogram image. The analysis has been done on the DDSM and MIAS dataset by using classification accuracy, specificity, and sensitivity as the parameter with three state of art algorithms i.e. SVM classifier (without any optimization technique), Firefly (SVM with Firefly optimization), ACO-PSO-GA (SVM with hybrid ACO-PSO-GA optimization). The improvement in the performance measures against three state of art techniques shows the significance of the algorithm.***

Index Terms: ***Mammogram, Classification Accuracy, Malignant tumor, ACO, PSO, GA, DE***

I. INTRODUCTION

Breast Cancer is one of the major healthcare issues with increasing death rate in women. The breast cancer can be dealt only if, it is detected at an early stage. The need for early detection process is fulfilled by the imaging technology [1]. Mammography is one of the imaging technique used to detect whether the tumor is benign or malignant (cancer tumor). The high classification accuracy is required for the tumor classification as the wrong result may take the life of a person [2]. A flowchart to describe the steps to detect breast cancer has been given in figure 1. The figure 1 shows that the input mammogram images are preprocessed to extract the region of interest from the image. Then the features specifically texture feature of the image are extracted which undergo a feature reduction step before the classification of the tumor [3]. In this process, feature reduction is one the most important step to select the significant features. This paper applies a meta-heuristic technique for the feature reduction as it requires optimization to select the significant features.

An optimization problem [4] involves finding the best solution from an exponentially large set of solutions. Mathematically an optimization problem $P(S|F)$ can be defined as follows:

Maximize $F(X)$ Subject to $gi(X) \leq 0, \quad i = 1,2, \dots \dots . . p$ and $p \geq 0$.

Where, S is the solution space, X is a vector of variables $X = \{X_1, X_2 \dots \dots \dots X_n\}$;$gi(x) \leq 0$ Is an inequality constraint.

In the feature reduction problems, one feature may not useful individually can be significant with other features so it is a multi-objective optimization problem [5].

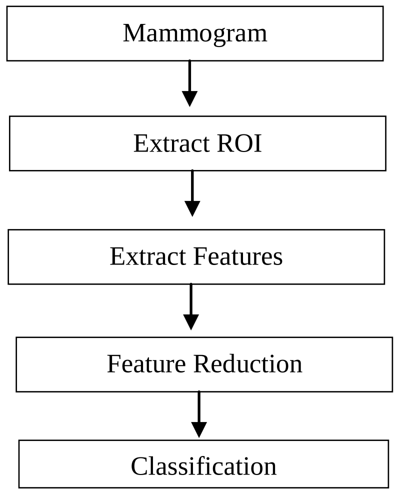

Figure 1: Malignant Tumor Detection Steps

Meta-heuristic techniques have become a better choice for solving the optimization problem. Metaheuristic word was the first time coined by Glover in 1986. These are higher level search strategies to guide the lower level heuristic to enhance their performance [6]. This approach provides a framework applicable to all the problems with the tuning of the parameter according to some specific requirement. Metaheuristic techniques possess some unique features for solving problems within a reasonable time frame such as to explore the search space efficiently to find an optimal solution, these techniques are independent of the type of problem and possesses a balance between intensification and diversification components to get stuck in local minima [7]. These techniques are very diverse such as ACO, PSO, GA, DE, etc. Each of these techniques has its own quality to solve the optimization problem. This paper discusses a hybrid meta-heuristic algorithm that combines the quality of diverse meta-heuristic techniques in the next section.

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Anju Bala*, Research Scholar, Department of Computer Science and Applications, M.D.U, Rohtak, India. E-mail: anjudeswal.mdu@gmail.com
Priti, Assistant Professor, Department of Computer Science and Applications, M.D.U, Rohtak, India. E-mail: pritish80@gmail.com

II. HYBRID ACO-PSO-GA-DE ALGORITHM

HAPGD Algorithm [8] removes the limitation of individual algorithms by combining the strengths of each algorithm. This algorithm combines exploration strength of ACO with exploitation property of PSO which are being balanced by the GA-DE [9][10]. The GA-DE crossover phase uses DE perturbation phase to improve the convergence. This algorithm removes the random initiation problem of GA by passing vectors processed by ACO and PSO [11]. HAPGD initiates with randomly generated vectors which are processed by the ACO and PSO for one iteration to explore and exploit the search space respectively. Then, the solution of the first iteration is passed to the GA which performs the Selection and mutation operation on vectors. Then the resultant vectors undergo the differential perturbation to generate input for the ACO and PSO [12][13]. This step is repeated until the algorithm converges. The complete details can be understood by the following algorithm:

A. Hybrid ACO-PSO-GA-DE Algorithm

1. Perform Initialization i.e. initiate ACO and PSO parameters and initiate number of iteration (nitr) to 0.
2. Initiate Random vector say H

$$H = [H_1 \quad H_2 \quad \cdots \quad H_n]$$

// this vector act as an initial vector for ACO as well as PSO

3. H_a=Apply ACO on H
4. Update Pheromone Value
5. H_p=Apply PSO on H
6. Update local and global path
7. H_a=QuickSort(H_a)
8. H_p=QuickSort(H_p)
9. H_1= $H_a(1)$
10. H_2= $H_p(1)$
11. F1=f(H_1)
12. F2=f(H_2)
13. While(nitr<MAX_ITERATION && change<e^{-10})
    - a. $U_1 = \frac{H_1 - LB_a}{UB_a - LB_a}$
    - b. $H_{1n} = H_1 + U_1 * (UB_a - LB_a)$
    - c. $U_2 = \frac{H_2 - LB_p}{UB_p - LB_p}$
    - d. $H_{2n} = H_2 + U_2 * (UB_p - LB_p)$
    - e. $r_1 = rand(1{:}n)$
    - f. $r_2 = rand(1{:}n)$
    - g. $H_{1n} = H_{1n} + rand * (H_{r1a} - H_{r2a})$
    - h. $H_{2n} = H_{2n} + rand * (H_{r1p} - H_{r2p})$
    - i. H_a=Apply ACO on H_{1n}
    - j. Update Pheromone Value
    - k. H_p=Apply PSO on H_{2n}
    - l. Update local and global path
    - m. H_a=QuickSort(H_a)
    - n. H_p=QuickSort(H_p)
    - o. H_1= $H_a(1)$
    - p. H_2= $H_p(1)$
    - q. Change=F1-f(H_1)+F2-f(H_2)
    - r. F1=f(H_1)
    - s. F2=f(H_2)
    - t. nitr++;

    End while
14. If(F1>F2)

        Return H1

    Else

        Return H2

    End if

HAPGD algorithm is capable to solve optimization problems efficiently. This algorithm starts with a random vector processed by ACO and PSO. The first iteration updates the H (random vector) to H_a by using ACO and H to H_p by using PSO algorithms. Then, the quick sort is applied to sort the components (vectors) of H_a and H_p based on their fitness values. Then the first i.e. best vector is selected as H_1 and H_2 from H_a and H_p respectively to be processed for other iterations. The fitness values calculated for these vectors are F1 and F2. The selection, mutation and differential perturbation is applied on these vectors to generate the H_{1n} and H_{2n} vectors which undergo ACO and PSO process respectively to generate H_a and H_p vectors respectively. The change in the fitness values of the updated and existing vectors is calculated. The process is repeated until the change minimizes or the number of iteration exceeded. The best vector from the resultant vector is the output of the algorithm. This process has been used in the breast cancer detection process discussed in the next section.

III. MAMMOGRAM CLASSIFICATION USING HAPGD ALGORITHM

The process of malignant tumor detection has been shown in the figure 2. The input mammogram image is preprocessed to extract the region of interest. The preprocessing step includes the two sub steps that are conversion of .pgm image to .jpg image, cropping of image to extract the ROI. Then the sixteen texture features are extracted from the gray scale image using the GLCM technique. GLCM stands for gray level co-occurrence matrix. It is basically a robust method to measure the texture feature of an image. The texture feature of any image depends upon the intensity distribution of any pixel corresponding to its position. The texture feature can be of first order, second order or higher order depending upon the pixel combinations used to calculate it. The GLCM techniques computes the first order and second order texture features. The first order texture features are computed by combining the values at an angle of 0^o,45^o,90^o,135^o to the current pixel. While the second order texture features like Contrast, Entropy, Difference Variance, are calculated by the first order matrix. Here, we have computed the first order as well as second order GLCM features for the given mammogram image. Then , The algorithm discussed in previous section i.e. hybrid ACO-PSO-GA-DE is applied foe the feature reduction. The **H**ybrid **A**CO-**P**SO-**G**A-**D**E (HAPGD) algorithm [8] selects the significant features then performs classification using SVM classifier. The classification process divides the dataset into K equal parts where the value of K depends upon the system configuration and the Dataset size. Then, the classification process using

SVM classifier is applied on the dataset K times. Each time one part is selected for the testing and remaining for the training purposes. The overall classification accuracy is the average of accuracy achieved in each case. This process also shown in figure 2 classify the mammograms effectively, the fitness function uses in the process is described in next subsection.

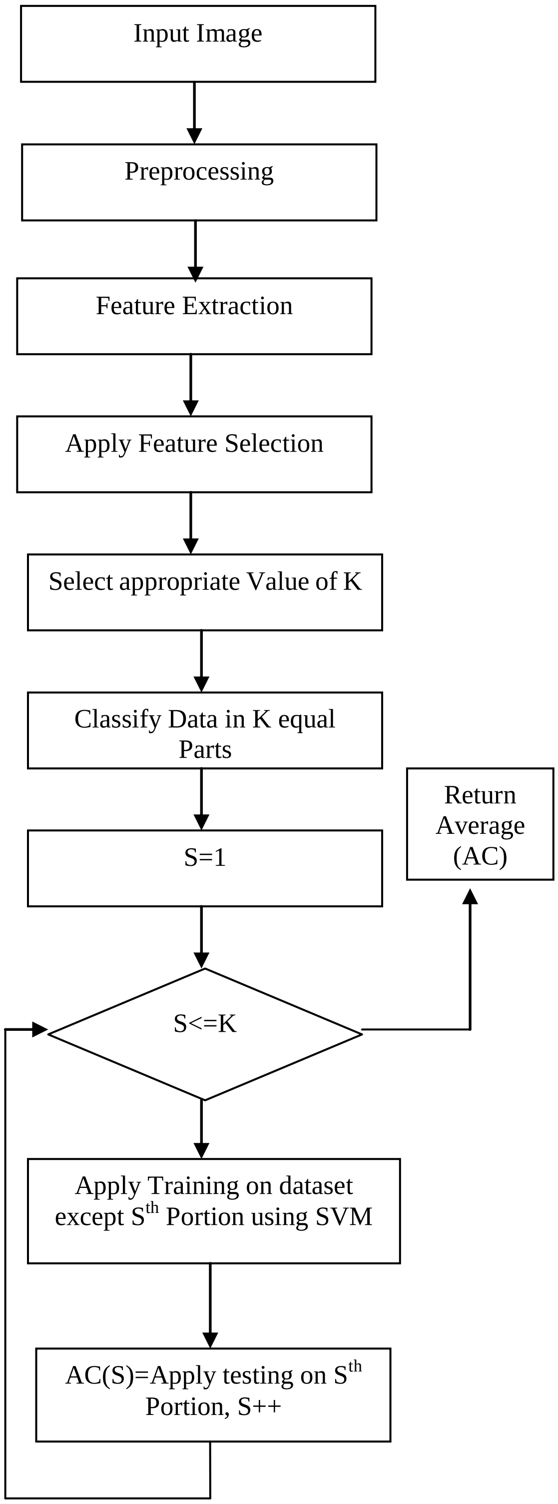

Fig 2: Mammogram Classification using HAPGD Classification Algorithm

A. Fitness Function

The mammogram classification using HAPGD algorithm has used following fitness function.

$$f = \theta_1 * senstivity + \theta_2 * accuracy$$

Here, $\theta_1 + \theta_2 = 1$. Here the values of $\theta_{1,}\theta_2$ are taken to be 0. 5, and 0. 5 respectively. The implementation and the result analysis of the process have been done in the next section.

IV. RESULT AND DISCUSSION

The algorithm has been analyzed on two mammograms datasets i.e. MIAS and DDSM available over the internet. The MIAS dataset is given by Mammographic Image Analysis Society which is an association of UK exploration assembles which utilizes 322 digitized images. The DDSM dataset is an acronym for Digital Database for Screening Mammography which is given by collective exertion between Massachusetts General Hospital, Sandia National Laboratories and the University of South Florida Computer Science and Engineering Department. The performance comparison has been done with three sate of art algorithms i.e. SVM classifier (without any optimization technique), Firefly (SVM with Firefly optimization [14]) [1], ACO-PSO-GA (SVM with hybrid ACO-PSO-GA optimization) [9] using sensitivity, specificity, and classification accuracy as the parameters. The analysis parameters i.e. accuracy, sensitivity and specificity described in next subsection.

A. Accuracy

Accuracy represents the number of instances correctly classified by the classifier. It is given as:

$$A = \frac{1}{n} * Correctly\_classified\_intsnaces \quad (1)$$

Here, n is the number of instances in the dataset

B. Sensitivity

It gives the correctly classified true instances. In other words, it is sensitivity to the correct classification. The sensitivity can be represented by (2).

$$Senstivity = \frac{TP}{TP+FN} \quad (2)$$

Here TP, FN shows true positive and false negative respectively.

C. Specificity

It is the correctly classified negative instances. It can be given by (3).

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

Here TN, FP denotes true negative and false positive respectively.

V. PERFORMANCE ANALYSIS

The performance of Mammogram classification using HAPGD algorithm is compared with SVM classifier (without any optimization technique), Firefly (SVM with Firefly optimization), ACO-PSO-GA (SVM with hybrid ACO-PSO-GA optimization) using the parameters described above on DDSM and MIAS datasets. The results are evaluated by executing the algorithms 10 times and taking the average of results. The initial parameter setting has been done as per the referred paper for each algorithm. The SVM classifier is used with RBF kernel having sigma=15. The classification accuracy is given in table 1.

Table 1: Classification Accuracy of Mammogram Classification using HAPGD algorithm

Dataset		SVM	Firefly	ACO +PSO +GA	HAPG D
		Accur acy	Accur acy	Accur acy	Accur acy
MIA S	NORMAL	0.9026	0.9224	0.9237	0.9423
	BENIGN	0.8416	0.838	0.841	0.8525
	MALGIN ANT	0.8511	0.8391	0.8472	0.9115
DDS M	NORMAL	0.9924	0.9955	0.9962	0.999
	BENIGN	0.9828	0.9838	0.984	0.9901
	MALGIN ANT	0.9178	0.9386	0.9439	0.9751

The comparison of classification accuracy on DDSM and MIAS datasets of HAPGD algorithm with other state or art algorithms is shown in table 1. The HAPGD algorithm exhibits more accuracy on each dataset as compared to other algorithms. This is due to balancing of exploration phase (due to ACO) and exploitation phase (PSO) by the GA. The comparison of the sensitivity is shown in the table 2.

Table 2: Sensitivity of Mammogram Classification using HAPGD algorithm

Dataset		SVM	Firef ly	ACO+PSO+ GA	HAPG D
		Se	Se	Se	Se
MIA S	NORM AL	0.82 42	0.833 2	0.8322	0.8542
	Benign	0.76 08	0.760 8	0.7608	0.7803
	Maligna nt	0.82 97	0.886 8	0.7725	0.9725
DDS M	NORM AL	0.97 56	0.981 2	0.9865	0.9923
	Benign	0.98 67	0.991 5	0.9915	0.9963
	Maligna nt	0.93 38	0.952 3	0.9781	0.9913

The comparison clearly shows the sensitivity of the HAPGD algorithm is better as compared to other hybrid techniques. This is due to efficient exploration and exploitation as ACO combined with the GA-DE gives the fast converging algorithm that explores the search space efficiently while the PSO combined with GA-DE results in algorithm that exploits the search space effectively. The comparison of the specificity on various datasets is shown in the table 3.

Table 3: Specificity of Mammogram Classification using HAPGD algorithm

Dataset		SVM	Firefl y	ACO + PSO +GA	HAPG D
		Sp	Sp	Sp	Sp
MIAS	NORM AL	0.9134	0.9224	0.9254	0.9532
	Benign	0.9437	0.9355	0.9355	0.9437
	Maligna nt	0.795	0.7338	0.8378	0.9195
DDS M	NORM AL	0.9887	0.99	0.9912	0.9934
	Benign	0.9703	0.9549	0.9549	0.9703
	Maligna nt	0.812	0.8544	0.7511	0.7511

The comparison in clearly denotes that the specificity of HAPGD algorithm is better as compared to other state of art algorithms due to more balanced exploration and exploitation search.

VI. CONCLUSION

This paper presents mammogram classification using HAPGD algorithm and measures the performance of the same on DDSM and MIAS datasets downloaded from internet. The comparison has been done with three states of art techniques including SVM classifier (without any optimization technique), Firefly (SVM with Firefly optimization), ACO-PSO-GA (SVM with hybrid ACO-PSO-GA optimization) by using classification accuracy, specificity and sensitivity as the parameters. The analysis clearly shows that the accuracy, as well as the sensitivity and specificity of the HAPGD algorithm, are better than the other algorithm due to balanced exploration and exploitation search. In the future, this work can be applied to other healthcare applications.

REFERENCES

1. Thawkar S, Ingolikar R. Classification of Masses in Digital Mammograms Using Firefly based Optimization. International Journal of Image, Graphics & Signal Processing. 2018 Feb 1;10(2).
2. Zhang BN, Cao XC, Chen JY, Chen J, Fu L, Hu XC, Jiang ZF, Li HY, Liao N, Liu DG, Tao O. Guidelines on the diagnosis and treatment of breast cancer (2011 edition). Gland surgery. 2012 May;1(1):39.
3. Wei J, Sahiner B, Hadjiiski LM, et al. Computer aided detection of breast masses on full field digital mammograms. Medical Physics 2005;32:2827–2838
4. R. Rajakumar, P. Dhavachelvan, and T. Vengattaraman, “A survey on nature inspired meta-heuristic algorithms with its domain specifications,” Proc. Int. Conf. Commun. Electron. Syst. ICCES 2016, 2016.
5. M. G. Torres, “Metaheuristics in Data Mining,” no. August, 2008.
6. R. Rao Kurada, K. K. Pavan, and A. D. Rao, “A Preliminary Survey on Optimized Multiobjective Metaheuristic Methods for Data Clustering Using Evolutionary Approaches,” Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 5, pp. 57–77, 2013.
7. X. S. Yang, S. Fong, X. He, S. Deb, and Y. Zhao, “Swarm Intelligence: Today and Tomorrow,” Proc. - 2016 3rd Int. Conf. Soft Comput. Mach.

Intell. ISCMI 2016, pp. 219–223, 2017.

8. Anju Bala, Dr. Priti, "Hybrid ACO-PSO-GA-DE Algorithm for Big Data Classification", has published in International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume - 8, Issue – 2, July 2019.
9. J. H. Tam, Z. C. Ong, Z. Ismail, B. C. Ang, and S. Y. Khoo, "A new hybrid GA−ACO−PSO algorithm for solving various engineering design problems," International Journal of Computer Mathematics, vol. 0, no. 0, pp. 1–37, 2018.
10. K. Menghour and L. Souici-Meslati, "Hybrid ACO-PSO based approaches for feature selection," International Journal of Intelligent Engineering and Systems, vol. 9, no. 3, pp. 65–79, 2016.
11. W. Y. Lin, "A GA-DE hybrid evolutionary algorithm for path synthesis of four-bar linkage," Mechanism and Machine Theory, vol. 45, no. 8, pp. 1096–1107, 2010.
12. S. Nemati, M. E. Basiri, N. Ghasem-Aghaee, and M. H. Aghdam, "A novel ACO-GA hybrid algorithm for feature selection in protein function prediction," Expert Systems with Applications, vol. 36, no. 10, pp. 12086–12094, 2009.
13. S. M.Uma, K. Rajiv Gandhi, E. Kirubakaran, and D. E. K. Dr.E.Kirubakaran, "A Hybrid PSO with Dynamic Inertia Weight and GA Approach for Discovering Classification Rule in Data Mining," International Journal of Computer Applications, vol. 40, no. 17, pp. 32–37, 2012.
14. H. Wang et al., "A hybrid multi-objective firefly algorithm for big data optimization," Applied Soft Computing Journal, vol. 69, pp. 806–815, 2018.

AUTHORS PROFILE

Anju Bala is MCA, Research Scholar (JRF), Department of Computer Science and Applications, M.D.U, Rohtak, India. She has published more than 10 publications in various journals of national /international repute.Her area of research includes Big Data and Data mining.

Dr. Priti, MCA, PhD (Computer Science), Assistant Professor**,** Department of Computer Science and Applications, M.D.U, Rohtak, India. She has published more than 50 publications in various journals/magzines of national and international repute. She has 12 years of experience in teaching and research. Her area of research includes Software Re-engg, Data Mining, Big Data, and Machine Learning.