Hybrid Optimized Framework for Classification of Breast Cancer

Ramani R.[1]* and Suthanthira Vanitha N.[2]
1. Department of Electronics and Communication Engineering, V.M.K.V. Engineering College, INDIA
2. Department of Electrical and Electronics Engineering, Knowledge Institute of Technology, Salem, INDIA
*ramani_gpt@yahoo.co.in

Abstract
Breast cancer is the second most common cancer after cervical cancer prevailing among women in India. Early breast cancer detection through regular mammography screening is an important factor in breast cancer survival as screen detected cancers are most likely to be diagnosed with more favorable prognostic factors than symptom detected cancers. Feature selection helps to reduce the feature space which improves the prediction accuracy and minimizes the computation time. The selected optimal features are considered for classification. Grey Wolf Optimization (GWO) is recently developed heuristics inspired from the leadership hierarchy and hunting mechanism of grey wolves in nature. In this work, hybrid Genetic Algorithm Grey Wolf Optimization (GA-GWO) and Artificial Bee Colony Grey Wolf Optimization (ABC-GWO) is proposed.

Keywords: Breast cancer, Grey Wolf Optimization (GWO), hybrid Genetic Algorithm Grey Wolf Optimization (GA-GWO) and Artificial Bee Colony Grey Wolf Optimization (ABC-GWO).

Introduction

Breast cancer is the most frequently diagnosed cancer amon g women and the second leading cause of cancer death amo ng women. Early detection of breast cancer increases treatment options and patients' survivability[1]. Although mammography is currently the most effective tool for early detection of breast cancer, it has some restrictions. On a screening mammographic examination, noncancerous lesions can be misinterpreted as a cancer (false-positive value), while cancers may be missed (false-negative value). As a result, radiologists fail to detect 10% to 30% of breast cancers. The false-positive value indicates the percentage of lesions that were found to be cancerous and subjected to biopsy. The miss rate in mammography is increased in dense breasts where the probability of cancer is four to six times higher than in non-dense breasts.

To enhance sensitivity of mammography, complimentary modalities such as ultrasound and Magnetic Resonance Imaging (MRI) are often recommended to achieve additional information. Recently, Computer-Aided Detection/Diagnosis (CAD) systems have been developed to reduce the expense and to improve the capability of radiologist in interpretation of medical images and differentiation between benign and malignant tissues[2]. The efficiency of radiologist's interpretation can be improved in terms of accuracy and consistency in detection/diagnosis, while his/her productivity can be improved by reducing the time required for reading the images. The computer outputs are derived using various techniques in computer vision to present some of the significant parameters such as the location of suspicious lesions and the likelihood of malignancy of detected lesions. Generally, CAD systems are executable on all imaging modalities and all kinds of examinations.

Texture helps to understand image content based on textural properties in images. Texture is the most important visual cue in identifying different types of homogeneous regions and gives information about the surface property, depth and orientation[3]. This texture information helps to extract specific characteristics from a data. Mammographic images possess textural information that could bear discriminant features. The Laws texture features were extracted from the mammogram to differentiate between abnormal and normal pixels.

Feature selection helps to reduce the feature space which improves the prediction accuracy and minimizes the computation time. This is achieved by removing irrelevant, redundant and noisy features .i.e., it selects the subset of features that can achieve the best performance in terms of accuracy and computation time. The selected optimal features are considered for classification. Till now no attempts have been made to hybrid the different feature selection algorithm to extract the feature from mammogram. Especially branch and bound techniques has been fully exploited to extract the feature from mammogram which is one of the best techniques to optimize the features among many features.

The advantages in using GAs are that they require no knowledge or gradient information about the response surface, they are resistant to becoming trapped in local optima and they can be employed for a wide variety of optimization problems. On the other hand GAs could have trouble in finding the exact global optimum and they require a large number of fitness functions evaluations. The advantage of using the honeybee search algorithm is the robustness against outliers. This algorithm searches the space through a coordinate and intelligent process that removes significantly the outliers. The proposed Artificial Bee Colony (ABC) optimization[4] algorithm can search for multiple thresholds which are very close to the optimal ones examined by the exhaustive search method.

Literature Survey

Mammography is a low dose x-ray procedure for the visualization of internal structure of breast. It detects about 80-90% of the breast cancers without any note of symptoms. A framework for classifying mammograms as tumor and no tumor is presented in Ramani & Vanitha[5]. Symlet wavelet and Singular Value Decomposition (SVD) were used for feature extraction and reduction respectively. Boosting algorithm was applied to predictive data mining to generate a sequence of classifiers. A hybrid learning Artificial Bee-AdaBoost (AB-AB algorithm) was proposed by combining concept of Artificial Bee Colony (ABC) algorithm and AdaBoost algorithm. The proposed hybrid algorithm boosts the classification ability of Support Vector Machine (SVM). MIAS dataset was used for evaluating the proposed method. Experimental results were conducted for AdaBoost and proposed optimization technique.

Otoom et al[6] made a review on types of feature sets and reviewed the implemented machine learning algorithms. Moreover, an extensive experiments was run to compare the classification performance of the aforementioned feature sets. Proposed results show that the image shape-based features are more discriminative for breast cancer classification when tested with ten-fold cross validation. To check the robustness of the best performing feature set, by examining with five-fold cross validation and with a variety of generative classification algorithms.

Karimi[7] proposed a new method for automatic detection of suspected breast cancer lesions using ultrasound is proposed. In this fully automated method, new de-noising and segmentation techniques were introduced and high accuracy classifier using combination of morphological and textural features was used. Used a combination of fuzzy logic and compounding to denoise ultrasound images and reduce shadows. A new method was introduced to identify the seed points and then use region growing method to perform segmentation. For preliminary classification three classifiers such as ANN, AdaBoost, FSVM were used and then used a majority voting to get the final result. Proposed method was demonstrated an automated system performs better than the other state-of-the-art systems. On proposed database containing ultrasound images for 80 patients reached an accuracy of 98.75% versus ABUS method with 88.75% accuracy and Hybrid Filtering method with 92.50% accuracy.

Jona & Nagaveni[8] proposed a CAD (Computer Aided Diagnosis) system to optimize the feature set using hybrid of Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) technique called Genetical Swarm Optimization (GSO) in Digital Mammogram. Even though PSO is a good optimization technique, it may be trapped in local minima and may prematurely converge. So, the genetic operators were used in PSO to overcome the difficulties. Feature selection plays a major role in diagnosis of mammogram. Gray Level Co-occurance Matrix (GLCM) texture features were extracted from the mammogram. All the extracted features do not help in detection of abnormality in a mammogram, so it is intended to reduce the feature set to improve classification accuracy. In this work, experiments are conducted on MiniMIAS database and Support Vector Machine (SVM) classifies the mammograms into normal and abnormal mammograms. Performance of GSO was compared with GA and PSO by means of Receiver Operating Characteristic (ROC) curve. Results show that, the GSO convergence performs better than both PSO and GA; GSO based SVM (GSO-SVM) classifier exhibits superior performance with an accuracy of 94% which is approximately 1% higher than GA based SVM (GA-SVM) and PSO based SVM (PSO-SVM) classification.

Methodology

MIAS Dataset: The images acquired by Mammographic Image Analysis Society (MIAS) dataset[9] are being used in the present study. The MIAS dataset were labeled and previously studied by expert radiologist having years of field experience. For the experimentation of the standard MIAS size of 322 images, only 305 images were considered. The 13 images with asymmetry in the MIAS dataset were not considered in this study and there were repetition of 4 images, so only one among the repeated two is being added, this leaves 305 images having size 1024x1024 pixels each. Figure 1 shows a sample image used for investigation.

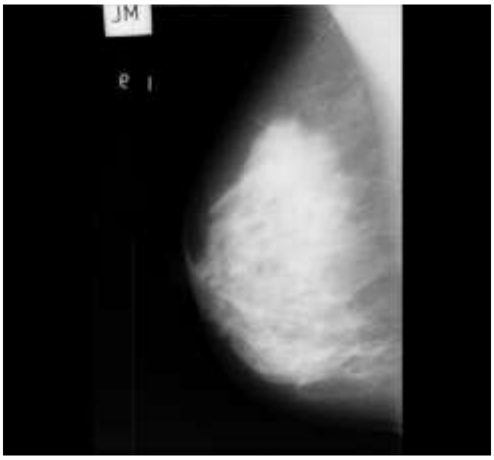

Figure 1: Sample Image of Breast with Tumor

Wavelet for Feature Extraction: Wavelet function gives signals multi-resolution representations, because all frequency components are examined with distinct resolutions and scales. This permits Wavelet Transforms to denote discontinuities in signals through short functions while concurrently highlighting low frequencies through wide functions.

Wavelets decompose signals f into sets of *scaling functions* through usage of wavelet function:

$$(W_a f)(b) = \int f(x)\psi^{*}_{a,b}(x)\,dx \quad (1)$$

Mother wavelet function $\psi(x)$ may be given as in (2):

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)$$
so that
$$\int \psi(x)\,dx = 0$$
Discrete wavelets are got through setting
$$a = 2^n$$
$$b \in \square \tag{2}$$

For colour images, wherein "time" refers to pixels spatial position while "frequency" refers to colour variations between consecutive pixels and basis function is as in equation (3):

$$\Psi_{j_1,k_1;j_2,k_2}(x_1,x_2) = \psi_{j_1,k_1}(x_1)\cdot\psi_{j_2,k_2}(x_2) \tag{3}$$

The initial transformation stage splits signals into four sub-images denoted by LL, LH, HL and HH filters. Next transformation level splits LL sub-image into four more sub-images as illustrated in figure 2.

<table>
<tr><td>$A^3_{2^{-3}} f$</td><td>$D^1_{2^{-3}} f$</td><td rowspan="2">$D^1_{2^{-2}} f$</td><td rowspan="3">$D^1_{2^{-1}} f$</td></tr>
<tr><td>$D^2_{2^{-3}} f$</td><td>$D^3_{2^{-3}} f$</td></tr>
<tr><td colspan="2">$D^2_{2^{-2}} f$</td><td>$D^3_{2^{-2}} f$</td></tr>
<tr><td colspan="3">$D^2_{2^{-1}} f$</td><td>$D^3_{2^{-1}} f$</td></tr>
</table>

Figure 2: Sub-images in the "wavelet" image representation

Feature Selection: The best feature subset is to be selected from the extracted features.

Singular Value Decomposition (SVD) for Feature Selection: Coefficients from symlets are decreased with SVD that decreases higher dimensional parameter datasets to low dimensional spaces disclosing initial data substructures evidently arranging it in descending order of variations. SVD makes sure that discovering initial datasets points' most optimal approximations with lesser dimensions. If A is symmetric as well as positive definite, orthogonal matrices like Q for which A = QΛQT is probable; wherein Λ denotes Eigen value matrices. SVD generates matrix A as product UΣVT wherein U as well as V are orthogonal while Σ denotes diagonal matrices wherein non-zero entries of Eigen values of ATA square roots[10].

Let X represent an $m \times n$ matrix of real-valued data and $rank^3 r$, wherein with no loss of generality $m \geq n$, hence $r \leq n$. Within microarray data, x_{ij} denotes expression levels of i^{th} gene in j^{th} array. Components of i^{th} row of X generate n-dimensional vectors g_i that are given as transcriptional responses of i^{th} gene. In an alternate fashion, components of j^{th} column of X generate m-dimensional vectors a_j that are given as expression profiles of j^{th} array.

Formula for SVD of X is in equation (4):
$$X = USV^T, \tag{4}$$

wherein U is $m \times n$ matrix, S is $n \times n$ diagonal matrix, while V^T is $n \times n$ matrix. Columns in U are known as left singular vector, $\{u_k\}$, generating an orthonormal basis for array expression profiles, such that $u_i \cdot u_j = 1$ for i = j, and $u_i \cdot u_j = 0$ else. Rows of V^T comprise components of right singular vector $\{v_k\}$, generating orthonormal basis for gene transcriptional response. The components of S are non-zero on diagonals, as well as beingknown as singular values.

Hence, $S = diag(s_1, \ldots, s_n)$. Moreover, $s_k > 0$ for $1 \leq k \leq r$, and $s_i = 0$ for $(r+1) \leq k \leq n$. By tradition, arranging singular vectors is defined by high-to-low ordering of singular values with greatest singular values in upper left index of S matrix. It is to be noted that for squares, symmetric matrix X, SVD is equal to diagonalization, or eigen value issue's solution.

A major outcome of singular value decomposition of X is given in equation (5):
$$X^{(l)} = \sum_{k=1}^{l} u_k s_k v_k^T \tag{5}$$

is the closest rank-l matrix to X. The term "closest" implies that $X^{(l)}$ brings to a minimum the sum of squares of difference of components of X and $X^{(l)}$, $\sum_{ij}\left|x_{ij} - x^{(l)}_{ij}\right|^2$. A method to compute Singular Value Decomposition is to initially compute V^T as well as S through diagonalizing $X^T X$:

$$X^T X = VS^2V^T, \tag{6}$$

Later for calculating U by (7)
$$U = XVS^{-1}, \tag{7}$$

wherein $(r+1), \ldots, n$ columns of V wherein $s_k = 0$ are not

AdaBoost Classifier: In 2003, Yoav Freund and Robert Schaphire formulated a machine learning meta-algorithm called Adaptive Boosting (AdaBoost). It is very simple to implement and good for generalization. It improves the classification accuracy and not prone to over fitting. It is an iterative algorithm and during an each iteration of the training phase, a new weak learner is added to create a strong learner that is only slightly correlated to the classifier. The weighting vector is adjusted every time the weak learner is added to the ensemble to focus on examples that were misclassified in the earlier iteration.

Hence it is called adaptive and finally results with a classifier with better accuracy. Adaptive boosting (AdaBoost) is a particular machine learning method used to train a series of weak classifiers. During the training process for AdaBoost, the weights of the training samples are adaptively updated after each boosting iteration. The weights of the training samples which are misclassified by the current component classifier are increased, while the weights of the training samples with correct classification are decreased. Finally, the weak classifiers are combined linearly to form a strong classifier[11].

For binary classifiers, AdaBoost may be delineated by (8):

$$H(z) = \mathrm{sgn}\left(\sum_{t=1}^{K} \delta_t h_t(z)\right)$$
$$\textit{wherein } \mathrm{K} \text{ is the quantity of weak learners}$$
$$\mathrm{h}_t \text{ refers to the learner}$$
$$\delta_t \text{ refers to the weight} \tag{8}$$

With every fresh iteration t, AdaBoost selects a fresh hypothesis h_t which classifies training examples with great classification errors in previous iterations.

$$h(z) = \begin{cases} 1 & \text{if } \ \mathrm{g}(f(z)) > \theta \\ -1 & \text{otherwise} \end{cases}$$
(9)

Grey Wolf Optimizer (GWO): Grey wolf (Canis lupus) belongs to Canidae family. Grey wolves are considered as apex predators, meaning that they are at the top of the food chain. Grey wolves mostly prefer to live in a pack. The group size is 5-12 on average. The leaders are a male and a female, called alphas. The alpha is mostly responsible for making decisions about hunting, sleeping place, time to wake, and so on. The alpha's decisions are dictated to the pack. The alpha wolf is also called the dominant wolf since his/her orders should be followed by the pack[12]. The alpha wolves are only allowed to mate in the pack. Interestingly, the alpha is not necessarily the strongest member of the pack but the best in terms of managing the pack. This shows that the organization and discipline of a pack is much more important than its strength.

The second level in the hierarchy of grey wolves is beta. The betas are subordinate wolves that help the alpha in decision-making or other pack activities. The beta wolf can be either male or female, and he/she is probably the best candidate to be the alpha in case one of the alpha wolves passes away or becomes very old. The beta wolf should respect the alpha, but commands the other lower-level wolves as well. It plays the role of an advisor to the alpha and discipliner for the pack. The beta reinforces the alpha's commands throughout the pack and gives feedback to the alpha.

The lowest ranking grey wolf is omega. The omega plays the role of scapegoat. Omega wolves always have to submit to all the other dominant wolves. They are the last wolves that are allowed to eat. It may seem the omega is not an important individual in the pack, but it has been observed that the whole pack face internal fighting and problems in case of losing the omega. If a wolf is not an alpha, beta, or omega, he/she is called subordinate (or delta in some references). Delta wolves have to submit to alphas and betas, but they dominate the omega. Scouts, sentinels, elders, hunters, and caretakers belong to this category. Scouts are responsible for watching the boundaries of the territory and warning the pack in case of any danger. Sentinels protect and guarantee the safety of the pack. Elders are the experienced wolves who used to be alpha or beta. Hunters help the alphas and betas when hunting prey and providing food for the pack. Finally, the caretakers are responsible for caring for the weak, ill, and wounded wolves in the pack.

In order to mathematically model the social hierarchy of wolves when designing GWO, considered the fittest solution as the alpha (α). Consequently, the second and third best solutions are named beta (β) and delta (δ) respectively. The rest of the candidate solutions are assumed to be omega (ω). In the GWO algorithm the hunting (optimization) is guided by α, β, and δ. The wolves follow these three wolves.

Grey wolves form circles around the prey in the course of their hunt. For a mathematical algorithm of the encircling activity, (10) is suggested:

$$\vec{D} = |\vec{C}.\vec{X}_P(t) - \vec{X}(t)|$$
$$\vec{X}(t+1) = \vec{X}_P(t) - \vec{A}.\vec{D} \tag{10}$$

wherein t denotes current iteration, $\vec{A}$ as well as $\vec{C}$ are coefficient vectors, $\vec{X}_P$ refers to position vectors of preys while $\vec{X}$ denotes position vectors of grey wolves.

Vectors $\vec{A}$ as well as $\vec{C}$ are computed through (11):

$$\vec{A} = 2\vec{a}.\vec{r}_1 - \vec{a}$$
$$\vec{C} = 2.\vec{r}_2 \tag{11}$$

Wherein components of $\vec{a}$ are linearly reduced from 2 to 0 during iterations while $\vec{r}_1, \vec{r}_2$ are arbitrary vectors in [0,1].

Grey wolves possess the capacity for recognition of position of prey and circle them. Hunts are typically influenced by alphas. Betas as well as deltas may also take part in hunts at times. But in abstract search spaces, there is no foreknowledge on position of prey, which is the optimum. For mathematical simulation of the hunting activity of grey wolves, it is assumed that alphas, which are the most optimal potential solutions, betas as well as deltas possess greater data regarding possible positions of prey. Hence, initial three best solutions are saved and other units are forced to keep the position updated as per the location of best search unit. The equation (12) below are introduced:

$$\vec{D}_\alpha = |\vec{C}_1.\vec{X}_\alpha - \vec{X}|, \vec{D}_\beta = |\vec{C}_2.\vec{X}_\beta - \vec{X}|, \vec{D}_\delta = |\vec{C}_3.\vec{X}_\delta - \vec{X}|$$
$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1.(\vec{D}_a), \vec{X}_2 = \vec{X}_\beta - \vec{A}_2.(\vec{D}_\beta), \vec{X}_2 = \vec{X}_\delta - \vec{A}_3.(\vec{D}_\delta)$$
$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_3 + \vec{X}_3}{3} \quad (12)$$

Grey wolves end the hunt through attacks on prey when they stop movement. For the mathematical model of the process of moving towards prey, value of $\vec{a}$ is reduced. It is to be noted that fluctuation range of $\vec{A}$ is also reduced by $\vec{a}$. Otherwise put, $\vec{A}$ is an arbitrary value ranging between [a, -a] wherein a is reduced from 2 to 0 during iterations. If arbitrary values of $\vec{A}$ are between [-1, 1], subsequent location of search agents may be in any location between current one and the location of prey.

Grey wolves typically explore with regard to locations of alphas, betas as well as deltas. They deviate from one another for searching for pretty and converge for the attacks. For mathematical model of this deviation, $\vec{A}$ is used with arbitrary values larger than 1 or lesser than -1 for obliging search agents to deviate from prey. This places emphasis on explorations and permits GWO for global searches. The pseudo code of GWO is:

Set the grey wolf population $X_i (i = 1, 2, ..., n)$
eSet a, A, and C Compute fitnesses of all search agents
X_a = best search unit
X_β = second best search unit
X_δ = third best search unit
while (t < Maximum quantity of iterations)
for all search units
classify using AdaBoost
Keep location of current search unit updated
end for
a, A, and C are to be updated
Compute fitness of every search agent
X_a, X_β and X_δ are to be updated
$t = t + 1$
end while
return X_a

Hybrid Genetic Algorithm- Grey Wolf Optimization (GA-GWO): The GA[13] starts with several alternative solutions to the optimization problem, which are considered as individuals in a population. These solutions are coded as binary strings, called chromosomes. The initial population is constructed randomly. These individuals are evaluated, using the partitioning-specific fitness function. This ensures that the GA can explore new features that may not be in the population yet. It makes the entire search space reachable, despite the finite population size.

Set the grey wolf population $X_i (i = 1, 2, ..., n)$, individual position
eSet a, A, and C Compute fitnesses of all search agents
X_a = best search unit
X_β = second best search unit
X_δ = third best search unit
Crossover method
a. Determine the best and worst individuals;
b. Improve the worst individual position using GA crossover;
*b.*1. *Select a crossover point*;
*b.*2. *Produce offspring* 1 *and offspring* 2;
*b.*3. *For each offspring* : *calculate fitness*;
*b.*4. *If any of the offspring are better than those existing in P*, *replace them*; *if not*, *generate a new random individual*;
while (t < Maximum quantity of iterations)
for all search units
classify using AdaBoost
Keep location of current search unit updated
end for
a, A, and C are to be updated
Compute fitness of every search agent
X_a, X_β and X_δ are to be updated
$t = t + 1$
end while
return X_a

Hybrid Artificial Bee Colony- Grey Wolf Optimization (ABC-GWO): In the ABC algorithm, each food source is considered as a possible solution to the problem to be optimized. The nectar amount represents the quality (fitness) of the solution represented by a food source. In the basic ABC algorithm, the initial population can be generated by using a random approach. Let $V_i = \{v_{i1}, v_{i2}, ...v_{in}\}$ represent the ith food source in the population, where n is the problem dimension. Each food source is generated as follows.

$$v_i^j = v_{\min}^j + rand[0,1](v^j_{\max} - v^j_{\min}) \quad (13)$$

Where $v^j_{\max}$ and $v^j_{\min}$ are the lower and upper bounds for the dimension j, respectively.

Set the grey wolf population $X_i (i = 1, 2, ..., n)$, initial population
eSet a, A, and C Compute fitnesses of all search agents
X_a = best search unit
X_β = second best search unit
X_δ = third best search unit
Calculate the initial cost function value, f(Sol);
Set best solution, Solbest Sol;
Set maximum number of iteration, NumOfIte;
Set the population size
do while (iteration < NumOfIte)
for i=1: EmployeedBee
Select a random solution and apply random neighborhood structure;
Sort the solutions in ascending order based on the Penalty cost;
Determine the probability for each solution, based on $p_i = \frac{\sum (1/fit_i)^{-1}}{fit_i}$
end for
for i=1: OnlookerBee
Sol* $\leftarrow$ select the solution who has the higher probability;
Sol** $\leftarrow$ Apply a random Nbs on Sol*;
if (Sol** < Solbest)
Solbest=Sol**;
end if
end for
Scoutbee determines the abandoned food source and replace it with the new food source.
iteration++
end do
while (t < Maximum quantity of iterations)
for all search units
classify using AdaBoost
Keep location of current search unit updated
end for
a, A, and C are to be updated
Compute fitness of every search agent
X_a, X_β and X_δ are to be updated
$t = t + 1$
end while
return X_a

Results and Discussion

The MIAS database contains ground truth of the image which includes the center of the mass and approximate radius of the circle enclosing the mass in terms of number of pixels. Table 1 shows the results of classification accuracy, precision, recall and F measures. Figure 3 to 6 shows the same.

It is observed from table 1 and figure 3 that the classification accuracy of ABC-GWO performs better by 10.3% than AdaBoost, by 6.9% than GA, by 3.7% than GWO, by 3.1% than ABC and by 1.2% than GA-GWO.

It is observed from table 1 and figure 4 that the precision of ABC-GWO performs better by 11.2% than AdaBoost, by 6.75% than GA, by 2.45% than GWO, by 3.52% than ABC and by 1.34% than GA-GWO.

It is observed from table 1 and figure 5 that the recall of ABC-GWO performs better by 12.39% than AdaBoost, by 8.9% than GA, by 4.6% than GWO, by 1.9% than ABC and by 1.34% than GA-GWO.

It is observed from table 1 and figure 6 that the F Measure of ABC-GWO performs better by 11.24% than AdaBoost, by 7.86% than GA, by 4.59% than GWO, by 2.56% than ABC and by 1.34% than GA-GWO.

Conclusion

Breast cancer is the most common form of cancer among women worldwide. Early detection of breast cancer can increase treatment options and patients' survivability. Mammography is the gold standard for breast imaging and cancer detection. However, due to some limitations of this modality such as low sensitivity especially in dense breasts, other modalities like ultrasound and magnetic resonance imaging are often suggested to achieve additional information. Recently, CAD systems have been developed to help radiologists in order to increase diagnosis accuracy. Results show that the classification accuracy of ABC-GWO performs better by 10.3% than AdaBoost, by 6.9% than GA, by 3.7% than GWO, by 3.1% than ABC and by 1.2% than GA-GWO.

References

1. Doi K., Computer-aided diagnosis in medical imaging: achievements and challenges, In World Congress on Medical Physics and Biomedical Engineering, Munich, Germany, Springer Berlin Heidelberg, 96-96 **(2009)**

2. Jalalian A., Mashohor S.B., Mahmud H.R., Saripan M.I.B., Ramli A.R.B. and Karasfi B., Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review, *Clinical Imaging*, **37(3)**, 420-426 **(2013)**

3. Dheeba J., Singh N.A. and Selvi S.T., Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach, *Journal of Biomedical Informatics*, **49**, 45-52 **(2014)**

4. Jegadeesan S. and Babu P., Mammogram Image Analysis for Micro-calcification Detection using Bee Colony Optimization, *International Journal of Computer Science And Technology*, *IJCST*, **3(3)**, 1012-1016 **(2012)**

Table 1
Results of Classification Accuracy, Precision, Recall and F measures

	AdaBoost	GA	GWO	ABC	GA-GWO	ABC-GWO
Classification Accuracy	0.88	0.91	0.94	0.9455	0.9636	0.9755
Precision	0.87	0.91	0.95	0.94	0.9607	0.97365
Recall	0.86	0.89	0.93	0.955	0.9607	0.97365
F measure	0.87	0.9	0.93	0.955	0.9607	0.97365

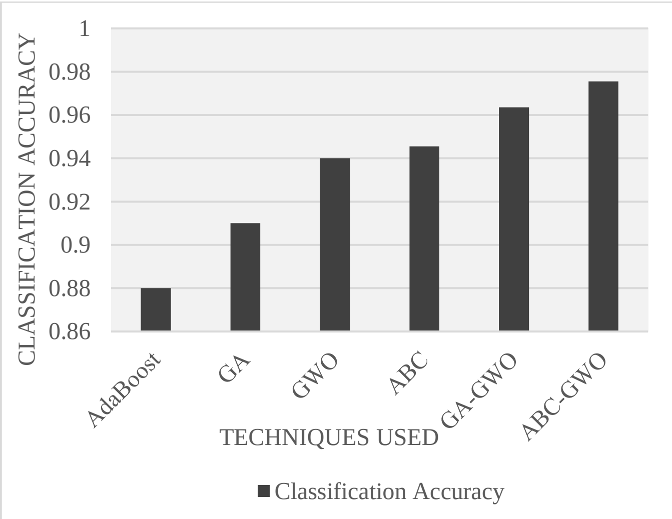

Figure 3: Classification Accuracy

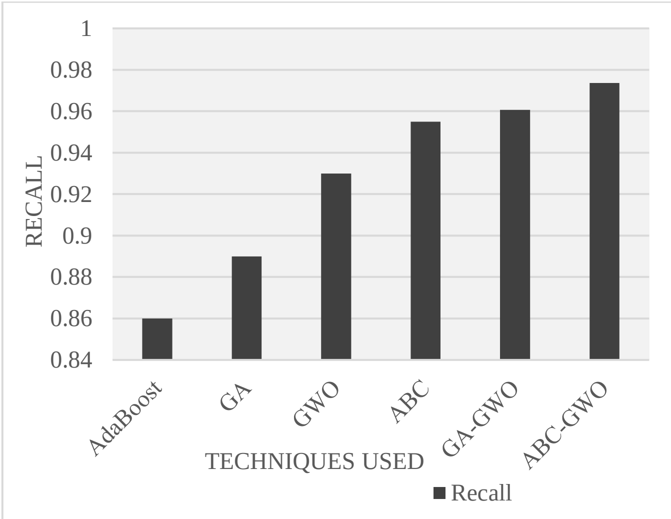

Figure 5: Recall

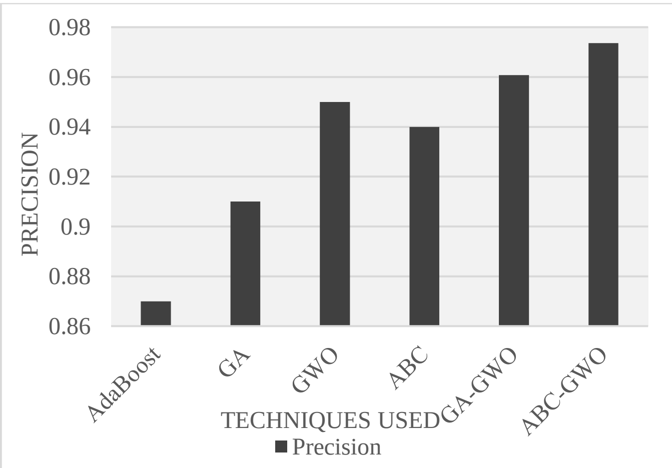

Figure 4: Precision

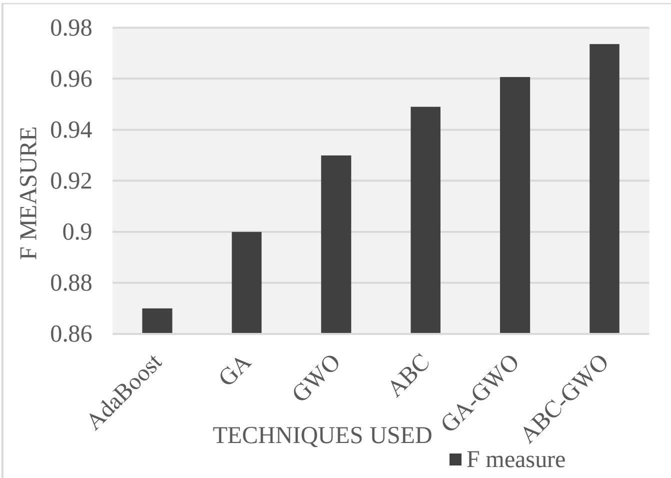

Figure 6: F Measure

5. Ramani R. and Vanitha N.S., Computer Aided Detection of Tumors in Mammograms Using Optimized Ada Boost, *Journal of Computational and Theoretical Nanoscience*, **13(8)**, 4982-4987 **(2016)**

6. Otoom A.F., Abdallah E.E. and Hammad M., Breast Cancer Classification: Comparative Performance Analysis of Image Shape-Based Features and Microarray Gene Expression Data, *International Journal of Bio-Science and Bio-Technology*, **7(2)**, 37-46 **(2015)**

7. Karimi B., An automatic system for classification of breast cancer lesions in ultrasound images, Doctoral dissertation, Concordia University **(2014)**

8. Jona, J. and Nagaveni N., A hybrid swarm optimization approach for feature set reduction in digital mammograms, *WSEAS Transactions on Information Science and Applications*, **9**, 340-349 **(2012)**

9. Gardezi S.J.S., Faye I. and Eltoukhy M.M., Analysis of mammogram images based on texture features of curvelet Sub-bands, In Fifth International Conference on Graphic and Image Processing, International Society for Optics and Photonics, 906924-906924 **(2014)**

10. Ramani R. and Vanitha N.S., Computer Aided Detection of Tumours in Mammograms, *International Journal of Image, Graphics and Signal Processing*, **6(4)**, 54 **(2014)**.

11. Vaidehi K. and Subashini T.S., Automatic Characterization of Benign and Malignant Masses in Mammography, *Procedia Computer Science*, **46**, 1762-1769 **(2015)**

12. Mirjalili S., Mirjalili S.M. and Lewis A., Grey wolf optimizer, *Advances in Engineering Software*, **69**, 46-61 **(2014)**

13. Mohanty A.K. and Lenka S.K., Efficient image mining technique for classification of mammograms to detect breast cancer, *Special Issue of IJCCT*, **2(2)**, 3 **(2010)**.