COMPUTER AIDED DETECTION OF TUMORS IN MAMMOGRAMS USING OPTIMIZED SUPPORT VECTOR MACHINES

Ramani R.¹, N. Suthanthira Vanitha²
¹Department of Electronics and Communication Engineering, V.M.K.V. Engineering College, India, ²Department of Electrical and Electronics Engineering, Knowledge Institute of Technology, Salem, India, E-Mail: ramanir.mam@gmail.com

ABSTRACT
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 this paper. Symlet wavelet and Singular Value Decomposition (SVD) are used for feature extraction and reduction respectively.Boosting algorithm is applied to predictive data mining to generate a sequence of classifiers. A hybrid learning Artificial Bee-AdaBoost (AB-AB algorithm) is 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 is used for evaluating the proposed method. Experimental results are conducted for AdaBoost and proposed optimization technique.

Keywords: mammography, MIAS dataset, Symlet wavelet, singular value decomposition (SVD), support vector machine (SVM), artificial bee colony (ABC), AdaBoost.

INTRODUCTION
Mammography is a low dose x-ray procedure used for the visualization of internal structure of breast and it has been proven as the most reliable method and a key screening tool for the early detection of breast cancer [1]. An X-ray beam is passed through the tissue for recording variations in amounts of radiation that are absorbed. Since different tissues in the breast will absorb different amounts of radiation, it is possible to distinguish features and information’s about the tissues examined. Generally, a digital mammogram detects the varying degrees of breast cancer as clustered micro calcifications, speculated lesions, circumscribed masses, ill-defined masses, and architectural distortions [2].

A mammogram has three kinds of tissues: breast supporting tissues (consisting of fibrous tissue and fat), lobes, and lesions (calcifications/masses) if any. There are two type of Mammography; “screening mammography” and “diagnostic mammography”. In a screening mammogram, each breast is X-rayed in two different positions as: from top to bottom and from side to side. When a mammogram image is viewed, the breast tissue appears as white and opaque and fatty tissue appears darker and translucent. A mammogram contains two regions specifically as: the exposed breast region and the unexposed non-breast region. It is necessary to identify the breast region first for the reduction of the processing and then to remove the non-exposed breast region [1].

Due to the implication of an automated image categorization helping the physicians and radiologists, most researches in the field of medical images classification are devoted to it. Since the classification algorithm needs data to be composed of feature vectors, data mining cannot be directly performed on the original image [3]. In mammogram classification, feature extraction is considered as the most effective step. Texture features are most commonly used in the analysis and interpretation of mammogram images [4].

Boosting algorithm is applied to predictive data mining to generate a sequence of classifiers. During deployment for prediction or classification of new cases, the predictions from the different classifiers can then be combined to derive a single best prediction or classification. Also, it can be applied to learning methods that do not explicitly support weights or misclassification costs. Boosting algorithms combine the weak learners to produce a complex decision boundary. Boosting iterations are gradient descent steps moves towards the predictor f(x) of minimum risk for the loss L[y, f(x)] =e⁻¹ (yf(x)) [5].

In this paper, the concept of artificial bee colony algorithm and adaptive boosting algorithm is combined and hybrid learning algorithm of AB-AB is proposed to boost the classification ability of Support Vector Machine (SVM). In section 2 literature review is discussed. Section 3 contains the methodology which includes MIAS dataset, Artificial Bee Colony (ABC) algorithm, Symlet. Sample mammogram image is taken and Symlet approximation, diagonal, vertical, horizontal based coefficients are calculated. Flowchart explains the overall methodology of this paper. Section 4 includes the results of classification accuracy, RMSE, Precision, Recall for AdaBoost and proposed optimization technique. In last section the paper is concluded.

LITERATURE SURVEY
Kilic et al. [7] proposed using Artificial Neural Network (ANN) with wavelets for the Mammographic Mass Detection. Results showed that the multilayer ANN with the Back propagation, Conjugate Gradient and Levenberg–Marquardt algorithms and ten-fold cross validation procedure was used where 89.2% was achieved with Levenberg–Marquardt algorithm. Haar Wavelet Features and Haralick texture features were used for mammogram
image classification by Joseph et al. [8]. An average classification accuracy of 98.6% for training, validation and testing was obtained. Differential Evolution Optimized Wavelet Neural Network (DEOWNN) for detection of tumor masses in mammograms was proposed by Dheeba et al. [9]. The proposed algorithm has a sensitivity of 96.9% and specificity of 92.9%. Raman et al. [10] presented region based segmentation using Haralick texture features for mammogram classification. Dheebaand Selvi [11] proposed using Swarm Optimization Neural Network (SONN) for detection of micro calcification in digital mammograms. The proposed method extracts the texture energy measures from Region of Interest (ROI) containing micro calcification. The neural network was optimally designed using Particle Swarm Optimization algorithm. Jasmine et al. [12] presented micro calcification detection based on wavelet analysis and neural networks, MIAS dataset was used to evaluate the proposed method.

Jaffer et al. [13] proposed using eight different multi-domain features to classify mammograms. Support vector machine (SVM) was used with Multilayer Perceptron’s (MLP) were used as classifiers with given 8 features. Results show the superiority of the proposed algorithm in terms of sensitivity, specificity and accuracy. Lahmri et al. [14] proposed a hybrid processing system using the discrete cosine transform and the Radon transform. Kim et al. [15] with the additional properties of margin-maximization and redundancy-minimization in order to further increase the accuracy. Sampalo et al. [16] proposed using cellular neural networks for segmentation, and the segmented regions were analyzed using shape descriptors and geostatic functions. SVM were used to classify the candidate regions as masses or non-masses with sensitivity 80%, rates of 0.84 false positives per image and 0.2 false negatives per image. Subashini et al. [17] developed a classification methodology with following steps: preprocessing, feature extraction and classification. Statistical features were extracted from region which signifies texture features of breast tissue and fed to SVM to classify into any of three classes: fatty, glandular and dense tissue. Multi-resolution representation of mammograms for extracting discriminative features was proposed by Eltoukhy et al. [18]. SVM was used to classify normal and abnormal tissues, distinguishes between benign and malignant tumors with high accuracy obtained [18].

Dehghan et al. [19] presented SVM classifiers with RBF kernel method and this was applied to a database of 40 mammograms (Nijmegen database) containing 105 clusters of MCs. Results of 89.55% mean true positive detection rate was achieved at the cost of 0.921 false positive per image. Fareeth et al. [20] proposed a continuous wavelet transform (1D - CWT) as feature selection technique and SVM as classifier and achieved excellent classification accuracy (100%) when compared with the other technique (1D - CWT and Fuzzy-C-mean clustering). Sharkas et al. [21] proposed discrete wavelet transforms (DWT), the contourlet transform, and the principal component analysis (PCA) for feature extraction, while the SVM was used for classification and achieved a classification rate was 100%. Yang et al. [22] proposed a virtual support vector SVM (VSVM) and the tangent vector SVM (TV-SVM). The experiment results showed that both techniques improve the performance in discriminating MCs from the image background, and TV-SVM achieved the best performance. In particular, the sensitivity was 96.3% for TV-SVM, compared to 94.5% for SVM, when the false positive rate was at 0.5%.

Moayedi et al. [23] proposed a kernel SVM is integrated with a Nero fuzzy rule-based classifier to form a support vector based fuzzy neural network (SVFNN) and attains classification accuracy of 96 % with efficient computation time for classifying mammograms. Multiple Support Vector Machine Recursive Feature Elimination (MSVM-RFE) was proposed by Yoon et al. [24] for feature selection. Yoon et al. [25] proposed AdaBoost based MSVM-RFE for classification of mammograms.

Land et al. [26] proposed both the Evolutionary Computation (EC)/Adaptive Boosting (AB) hybrid and SVM as pattern classifiers which results with best SVM configurations for optimum specificity and positive predictive value at very high sensitivities. Tirtajaya et al. [27] proposed dual-tree complex wavelet transform (DT CWT) as feature extraction and SVM as classifier and results demonstrate that a good classification accuracy was achieved. Ren et al. [28] proposed a balanced learning with optimized decision making to evaluate the performance of ANN and SVM where ANN outperforms the SVM when balanced learning was absent and performance of two classifiers becomes very comparable when both balanced learning and optimized decision making were employed.

Elsayad et al. [29] evaluates two Bayesian network classifiers: Naïve Bayesian and Markov blanket estimation on the prediction of severity of breast masses. The prediction accuracies of Bayesian networks were benchmarked against the multilayer perceptron neural network. Prathibha et al. [30] proposed DWT feature and was merged with DCT features. Classification was done with a combination of nearest neighbor classifiers; kNN, class based nearest neighbor and density based nearest neighbor. Eltoukhy et al [31] proposed an automatic detection of masses in digital mammograms which utilizes correlation between mass region and mammogram image to determine and extract the suspicious region in tested image. Evaluation done with 116 mammogram images from MIAS Dataset and achieves classification accuracy as 94.66 %.

Chen et al. [32] proposed method for automatic identification of breast boundary based on segmentation approaches. Results were obtained for breast-background segmentation, 98.8% and 91.5% accurate and for pectoral muscle segmentation, 92.8% and 87.9% accurate were achieved using datasets MIAS and EPIC.

Shelda et al. [33] presented the enhancement parameter of Contrast Limited Adaptive Histogram Equalization (CLAHE) based on Local Contrast
Modification (LCM), the wavelet transform was used to extract image features and results were tested on MIAS database and the accuracy of 93.41 percent was obtained. Javadi et al. [34] used PSO for feature selection. Fuzzy rules were formed and based on the rules the mammograms were classified with improved specificity and sensitivity. Genetic algorithm (GA) and New Particle Swarm Optimization (NPSO) for feature selection was proposed by Geetha et al. [35] and the selected features were classified using a three-layer Back propagation Network hybridized using PSO [35]. Karnan et al. [36] analyzed with GA and Ant Colony Optimization (ACO) for mammogram classification of micro calcification. Features were fed to a three-layer back propagation network hybridized with ACO for classification. [36]. Suganthi et al. [37] proposed multilayer neural network with ACO and PSO for classification of mammograms. Multi-objective GA was used for extracting optimal feature set.

MATERIALS AND METHODOLOGY

Context Figure-1 shows the framework of the mammogram classification method proposed in this paper. Following sections detail the methods used in the proposed framework.

Dataset

The Mammography Image Analysis Society (MIAS) is an organization for the research groups interested in understanding the mammograms, and a digital mammography database has been produced [38]. MIAS database consists of digitized images to 50 micron pixel edge with a Joyce-Loebl scanning microdensitometer. A linear device can be represented with a 0-3.2 optical density range pixels in an 8-bit word [39]. The database consists of 322 digitized films and radiologist “truth”-markings on a detected abnormalities location and is reduced to 200 micron pixel edge with padding/clipping to ensure all images were 1024 x1024 pixels at 8 bits per pixel. Erosion followed by dilatation with same structuring element, completes the opening function. The Mini MIAS database excludes the excessive network training and a better system generalization. Figure-2 shows a sample image used for investigation.

Figure-2. Sample image.

Symlet wavelets

The symlets are nearly symmetrical, orthogonal and bi-orthogonal wavelets proposed by Daubechies as modifications to the db family. In symN, N means the order. Some will use 2N instead of N. Symlets when applied to signal will performs a better and SNR of reconstructed or denoised signal is improved [40]. Symlets are Daubechies’ approximately a symmetry wavelets and these are orthogonal wavelets with close to symmetric scaling function [41]. Symlets are nearly symmetrical wavelets which is created to modify the Daubechies (db) wavelet family with properties of both wavelet families being similar.

\[
h(z) = \sum_k h_k z^{-k} \quad \text{and} \quad g(z) = zh\left(-z^{-1}\right)\]

where \(h\) and \(g\) are wavelet decomposition (analysis) filters, with \(h\) being a low pass filter and \(g\) is a high pass filter.

When \(A\) is symmetric and positive definite, an orthogonal matrix \(Q\) for which \(A = Q\Lambda Q^T\) is possible. where \(\Lambda\) is an Eigen values matrix. Singular Value Decomposition (SVD) formulates matrix \(A\) as a product \(U \Sigma V^T\) where \(U\) and \(V\) are orthogonal and \(\Sigma\) is a diagonal matrix where non-zero entries of Eigen values of \(A^TA\) square roots. The \(U\) and \(V\) columns provide bases for 4 fundamental subspaces [42]. Coefficients from Symlet can be reduced with SVD.
Value decomposition (SVD)

Singular Value Decomposition (SVD) reduces a high dimensional, variable data set to a lower dimensional space exposing original data substructure clearly ordering it from the highest variation to the least. SVD ensures finding original data sets points’ best approximation with fewer dimensions.

Definition of SVD

Any real $m \times n$ matrix $A$ can be decomposed uniquely as $A = U \Sigma V^T$,

$U$ is $m \times n$ and column orthogonal (its columns are eigenvectors of $A A^T$)

$V$ is $n \times n$ and orthogonal (its columns are eigenvectors of $A^T A$)

$\Sigma$ is $m \times n$ diagonal (non-negative real values called singular values)

$D = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_n)$ ordered so that $(\sigma_1 > \sigma_2 > ... > \sigma_n)$

If $U = (u_1, u_2, ..., u_n)$ and $V = (v_1, v_2, ..., v_n)$, then

$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T$

(Actually, the sum goes from 1 to $r$ where $r$ is the rank of $A$)

Adaptive Boosting (AdaBoost)

Adaptive Boosting (AdaBoost) is the popular ensemble method to enhance prediction accuracy of the base learner. Multiple classifiers are generated with this AdaBoost learning algorithm to utilize them to build as a best classifier. This requires less user knowledge for computing for improving accuracy over data sets [6]. Also it is used for maintaining a set of weights over the training set. The training set $(x_1, y_1), ..., (x_n, y_n)$ where each $x_i$ belongs to instance space $X$ and each $y_i$ is in the label set $Y = \{-1, +1\}$. The steps for AdaBoost are as follows:

1. Assign $N$ example $(x_1, y_1), ..., (x_n, y_n); x_i \in \{-1, +1\}$
2. Initialize the weights of $D(i) = 1 / N; i = 1, ..., N$
3. For $k = 1, ..., K$
4. Train weak learner using distribution $D_k$
5. Get weak hypothesis $h_k : X \rightarrow R$ with its error:

\[ \varepsilon_k = \sum_{i=\text{sign}(h_k(x), y)} D_k(i) \]

(1)

6. Choose $\varepsilon_k = R$

(2)

7. Update $D_k(i) = D_k(i) \exp(-\alpha_k y_k h_k(x)) / \tilde{z}_k$

(3)

where $\tilde{z}_k$ is the normalization factor.

8. Output the final hypothesis:

\[ H(x) = \text{sign} \left( \sum_{k=1}^{K} \alpha_k h_k(x) \right) \]

(4)

Artificial bee colony optimization

The ABC algorithm is developed by scrutinizing the behaviors of the real bees on finding the food source, which is called as the nectar, and shares the information of food sources to the bees in the nest [43]. Also in ABC, the artificial agents are defined and are classified into three types: the employed bee, the onlooker bee, and the scout. Each of type plays different role in the process such as

- The employed bee will stay on a food source and provides the neighborhood of the source in its memory.
- The onlooker will get the information of food sources from the employed bees in the hive and select any one of the food source to gather the nectar.
- The scout is responsible for finding the new food, the new nectar and sources [44].

The main goal of bees in ABC model is to find the best solution, the position of a food source represents a possible solution for the optimization problem and the nectar amount of a food source corresponds to the fitness of the associated solution [45].

Steps of ABC algorithm are given below:

- To initialize the food source positions.
- Each employed bee produces a new food source in her food source site and exploits the better source.
- Each onlooker bee selects a source depending on the quality of her solution, produces a new food source in selected food source site and exploits the better source.
- To determine the source to be abandoned and allocate its employed bee as scout for searching new food sources.
- To memorize the best food source found so far.

Repeat steps 2-5 until the stopping criterion is met [46].

Proposed hybrid learning artificial Bee - AdaBoost (AB-AB algorithm)

In the proposed hybrid learning Artificial Bee - AdaBoost (AB-AB algorithm), the ABC evolves and selects features and the AdaBoost bases its classifiers using the selected features. The advantage of the proposed algorithm is that the computational cost of the AdaBoost is lowered considerably.

The steps for proposed algorithm are as follows:

1. Assign $N$ example $(x_1, y_1), ..., (x_n, y_n); x_i \in \{-1, +1\}$
2. Initialize the weights of $D(i) = 1 / N; i = 1, ..., N$
3. For $k = 1, \ldots, K$
4. Train weak learner using distribution $D_k$
5. Initialization of Artificial Bee Colony to optimize weak classifier

Initialize the initial population and Evaluate fitness;
Calculate the initial cost function value, $f(Sol)$;
Set best solution, $Sol_{best} \leftarrow Sol$;
Set maximum number of iteration, NumOfIt;
Set the population size;
//where population size = OnlookerBee = EmployeeedBee;
iteration $\leftarrow 0$;
do while (iteration $< NumOfIte$)
   for $i=1$: EmployeeedBee
      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 the following formula:
      $$P_i = \frac{1}{\sum (U_{fit})^{-1}}$$
   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 ($Sol^{**} < Sol_{best}$) $Sol_{best}=Sol^{**}$;
   end if
   Scout bee determines the abandoned food source and replace it with the new food source.
   iteration++
end do
6. Evaluate hypothesis $h_k : X \rightarrow R$ with its error:
   $$\varepsilon_k = \sum_{i=1}^{n} D_k(i)$$
7. Choose $\varepsilon_k = R$
8. Update
   $$D_k(i) = \frac{D_k(i) \exp(-\alpha_k y h_k(x_i))}{z_k}$$
   where $z_k$ is the normalization factor.
9. Output the final hypothesis:
   $$H(x) = \text{sign} \sum_{k=1}^{K} \alpha_k h_k(x)$$

RESULTS AND DISCUSSIONS
The images for MIAS dataset are used for evaluating the proposed method. Features from the mammograms are extracted and selected using Symlet wavelets and SVD. The features are classified using AdaBoost and the proposed method. Table-1 shows the Classification Accuracy and Root Mean Square Error for AdaBoost and proposed optimization technique.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Classification accuracy</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ada Boost</td>
<td>87.88</td>
<td>0.2314</td>
</tr>
<tr>
<td>proposed optimization technique</td>
<td>94.55</td>
<td>0.1794</td>
</tr>
</tbody>
</table>

The images for MIAS dataset are used for evaluating the proposed method. The proposed optimization technique achieves a better performance by 7.59% compared to the AdaBoost technique.

It is observed from Figure-3 that Classification Accuracy is analyzed between techniques. The proposed optimization technique achieves a better performance by 7.59% compared to the AdaBoost technique.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Number of Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ada Boost</td>
<td></td>
</tr>
<tr>
<td>Proposed optimization technique</td>
<td></td>
</tr>
</tbody>
</table>

It is observed from Figure-4 that RMSE is analyzed between techniques. The proposed optimization technique decreases by 22.47% compared to the AdaBoost technique. The Table-2 shows the Precision and...
Recall table for AdaBoost and proposed optimization technique.

**Table-2. Precision and recall**

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ada Boost</td>
<td>0.8618</td>
<td>0.868182</td>
</tr>
<tr>
<td>Proposed optimization technique</td>
<td>0.9404</td>
<td>0.936364</td>
</tr>
</tbody>
</table>

**Figure-5. Precision.**

It is observed from Figure-5 that Precision is analyzed between techniques. The proposed optimization technique achieves a better performance by 9.12% when compared with AdaBoost technique.

**Figure-6. Recall.**

It is observed from Figure-6 that Recall is analyzed between techniques. The proposed optimization technique achieves a better performance by 7.85% when compared with AdaBoost technique.

**Figure-7. Performance of ABC algorithm.**

From the Figure-7 it is observed that the performance of Artificial Bee Colony (ABC) algorithm is performed. The best fitness value for the number of iterations is performed. Also, it is shown that from iteration= 137 till 150, the best fitness is convergence.

**CONCLUSIONS**

Mammography is a low dose x-ray procedure for the visualization of internal structure of breast. Symlet wavelet is used for feature extraction from the mammograms. Singular Value Decomposition is used for feature reduction. A hybrid learning Artificial Bee - AdaBoost (AB-AB algorithm) is proposed to boost the classification ability of Support Vector Machine (SVM). Results are compared with AdaBoost and Proposed optimization technique for classification accuracy, RMSE, precision and recall. Performance of ABC algorithm is presented for best fitness value.
REFERENCES


