ISSN: 0976-3104
SPECIAL ISSUE: Emerging Technologies in Networking and Security (ETNS)
Devisuganya and Suganthe

ARTICLE OPEN ACCESS

OPTIMIZED FEATURE SELECTION FOR BREAST CANCER DETECTION

S Devisuganya[1]*, RC Suganthe[2]

[1]*Velalar College of Engineering and Technology, Erode, Tamilnadu, INDIA*
[2]*Kongu Engineering College, Tamilnadu, INDIA*

ABSTRACT

Breast cancer is a common life-threatening cancer affecting woman. Mammography is an effective screening tool; radiologists use for breast cancer detection. The major phase in diagnosing breast cancers is features extraction and selection. Detecting tumours naturally requires extraction of features as well as their classification. This work presents a mammogram framework for detection of cancer. Pseudo Zernike Moments and Gaussian Markov Random Field (GMRF) are used for feature extraction. To reduce the dimensionality of the feature set, a Hybrid Shuffled frog-PSO algorithm is proposed. The work focuses on improving classification performance through feature selection. Experimental results demonstrate the effectiveness of the proposed method in improving the classification of the mammograms

Published on: 2nd -December-2016

KEY WORDS

Pseudo Zernike Moments, Gaussian Markov Random Field, Shuffle frog algorithm, Particle Swarm Optimization

***Corresponding author: Email:** devisuganya.cse@gmail.com

COMPUTER SCIENCE

INTRODUCTION

Alzheimer's disease Breast cancer is most common cancer among women. Mammographic images are X-ray images of breast region. The commonly used diagnostic technique including biopsy, mammography, thermography and ultrasound image. Among these techniques mammography is best approach for early detection. In early stage visual clues are subtle and varied in appearance, it makes diagnosis difficult. The abnormalities are hiding by breast tissue structure. Breast cancer detection and classification of mammogram images is the standard clinical practice for the diagnosis of breast cancer.

Mammography is the efficient tool available for the detection of breast cancer before physical symptoms appear. The earlier the cancer detection is challenging and difficult task. The biopsy is a standard approach for cancer detection manually under a microscope. But biopsy is difficult and time consuming task [1]. Breast cancer is considered a major health problem in western countries. A recent study from the National Cancer Institute (NCI) estimates that, in the United States, about 1 in 10 women will develop breast cancer during their lifetime. Moreover, in such country, breast cancer remains the leading cause of death for women in their 40s.

Although manual screening of mammographies remains the key screening tool for the detection of breast abnormalities, it is widely accepted that automated Computer Aided Diagnosis (CAD) systems are starting to play an important role in modern medical practices [2]. Early detection of breast cancer increases the survival rate and increases the treatment options. Screening mammography, x-ray imaging of the breast, is currently the most effective tool for early detection of breast cancer. Screening mammographic examinations are performed on asymptomatic woman to detect early, clinically unsuspected breast cancer [3]. Early detection via mammography increases breast cancer treatment options and the survival rate. However, mammography is not perfect.

Detection of suspicious abnormalities is a repetitive and fatiguing task. Screening mammography is widely used for early detection of breast cancer. Biopsy is invasive procedure and makes patient discomfort. Digital mammography is proven as efficient tool to detect breast cancer before clinical symptoms appear. Digital mammography is currently considered as standard procedure for breast cancer diagnosis. Various artificial intelligence techniques such as artificial neural network and fuzzy logic are used for classification problems in the area of medical diagnosis [4].

Feature selection is an important in breast cancer detection and classification. After features extraction, not all features are used to differentiate between normal and abnormal patterns. The advantage of limiting input features to make accuracy and reduce computation complexity. Many features are extracted from digital mammograms, they include region-based features, shape-based features, texture based features selection, and position based features. Texture features classify normal and abnormal in digital mammogram patterns. Feature classifies masses as benign or malignant using selected features. There are various methods used for mass classifications and some popular techniques are artificial neural networks and linear discriminating analysis [5].

Feature extraction is the first step in breast cancer detection. Texture feature is important for image classification. Various techniques have been used for computing texture features [1]. Grey-Level Co-Occurrence Matrices (GLCMs) is a powerful tool for image feature extraction. Gray level pixel distribution described by statistics like probability of two pixels having particular gray level at particular spatial relationships. This spatial information is provided as two dimensional gray level matrices. Image feature extraction is important step in mammogram classification. These features are extracted using image processing techniques. Several features are extracted from digital mammograms including texture feature, position feature and shape feature etc. [4].

Most of the limitations of conventional mammography can be overcome by using digital image processing. Thus, in order to improve the correct diagnosis rate of cancer, image enhancement techniques are often used to enhance the mammogram and assist radiologists in detecting it. Some of the efficient enhancement algorithm of digital mammograms based on wavelet analysis and modified mathematical morphology. Adopt wavelet-based level dependent thresholding algorithm and modified mathematical morphology algorithm to increase the contrast in mammograms to ease extraction of suspicious regions known as Regions of Interest (ROIs) are used [6].

Some scholars applied data mining techniques to predict diagnossis for digital mammography. Data mining techniques offer precise, accurate, and fast algorithms for such classification using dimensionality reduction, feature extraction, and classification routines. Neural networks have improved accuracy rate for the classification of benign and malignant patterns in digitized mammography. Feature selection is also commonly used in machine learning. It has already seen application in statistics, pattern recognition, and data mining. The aim of feature selection is to filter out redundant or irrelevant features from the original data.

Feature selection, a pre-processing step in the data mining process, is the step to select and extract more valuable information in massive related materials. It can improve the model’s performance as well as reduce the effort of training the model [7]. Feature selection is a main point that should be taken under consideration when implementing a CADx system for recognizing breast tissue. Selecting the most significant features that have the capability to describe and maximize the differences between different tissues in an ample way. Feature selection is an important factor that directly affects the classification result.

Most systems extract features to detect abnormalities and classify them as benign or malignant. The classification of malignant and benign is still a challenging problem for researchers. There are various feature extraction methods that serve to condense input data and to reduce redundancies by highlighting important characteristics of the image. The features of digital images can be extracted directly from the spatial data or from a different space after using a transform such as Fourier transform, wavelet transform or curvelet transform [8].

This study proposes feature selection based on shuffled frog and PSO. The remaining sections organized as: Section 2 reviews the related work in literature. Section 3 explains the methods which are used in the proposed work. Section 4 discusses the experiment results and section 5 concludes the proposed work.

RELATED WORK

Rehman et al., [9] proposed diverse features based breast cancer detection (DF-BrCanD) system to detect breast cancer that may be considered as a second opinion. The authors have used phylogenetic trees, statistical features and local binary patterns to generate a set of diverse and discriminative features for subsequent classification. Finally, Support Vector Machine with RBF kernel is used for the classification of mammographic images as cancerous and non-cancerous. The performance of the proposed DF-BrCanD system is analyzed using standard database for screening mammography through experimental comparison based on various performance measures. The authors showed that the proposed DF-BrCanD system is quite effective in detecting breast carcinoma.

Patel and Sinha [10] introduced a novel approach for accomplishing mammographic feature analysis through detection of tumor, in terms of their size and shape with experimental work for early breast tumor detection. The objective is to detect the abnormal tumor/tissue inside breast tissues using three stages: Pre-processing, Segmentation and post processing stage. By using pre-processing noise are remove and then segmentation is applied to detect the mass, after that post processing is applied to find out the benign and malignant tissue with the affected area in the cancers breast image. Size of tumor is also detected in these steps. The occurrences of cancer nodules are identified clearly.

Ganesan et al., [11] presented a one-class classification pipeline for the classification of breast cancer images into benign and malignant classes. Because of the sparse distribution of abnormal mammograms, the two-class classification problem is reduced to a one-class outlier identification problem. Trace transform, which is a generalization of the Radon transform, has been used to extract the features. Several new functional specific to mammographic image analysis have been developed and implemented to yield clinically significant features. Classifiers such as the linear discriminant classifier, quadratic discriminant classifier, nearest mean classifier, support vector machine, and the Gaussian Mixture Model (GMM) were used.

Deshpande et al., [12] made an attempt to build classification system for mammograms using association rule mining based on texture features. The proposed system used most relevant GLCM based texture features of mammograms. New method was proposed to form associations among different texture features by judging the importance of different features. Resultant associations can be used for classification of mammograms. Experiments were carried out using MIAS Image Database. The performance of the proposed method was compared with standard Apriori algorithm. The authors also investigated the use of association rules in the field of medical image analysis for the problem of mammogram classification.

Sanae et al., [13] presented an efficient classification of mammograms using feature extraction. In this approach the authors proposed to use comprehensive statistical Block-Based features, derived from all sub-bands of Discrete Wavelet decomposition. The classification of these features was performed using the Support Vector Machine (SVM). The evaluation of the proposed method was applied on Digital Database For Screening Mammography (DDSM). The system classifies normal from abnormal cases with high accuracy rate (96%). Comparative experiments have been conducted to evaluate the proposed method.

Kim [14] proposed a new classification technique that is based on support vector machines with the additional properties of margin-maximization and redundancy-minimization in order to further increase the accuracy. The author have conducted experiments on publicly available data set of mammograms and the empirical results indicated that the proposed technique performed superior to other previously proposed support vector machines-based techniques.

Thangavel and Velayutham [15] proposed a novel unsupervised feature selection method using rough set based entropy measures. A typical mammogram image processing system generally consists of image acquisition, pre-processing, segmentation, feature extraction and selection, and classification. The proposed unsupervised feature selection method was compared with different supervised feature selection methods and evaluated with fuzzy c-means clustering in order to prove the efficiency in the domain of mammogram image classification.

Aroquiaraj and Thangavel [16] proposed a novel unsupervised feature selection in mammogram image, using tolerance rough set based relative reduct. And also, compared with Tolerance Quick Reduct and particle swarm optimization (PSO) - Relative Reduct unsupervised feature selection methods. A typical mammogram image processing system generally consists of mammogram image acquisition, pre-processing of image segmentation, feature extraction, feature selection and classification. The proposed method is used to reduce features from the extracted features and the method is compared with existing unsupervised features selection methods. The proposed method is evaluated through clustering and classification algorithms in K-means and WEKA.

Wong et al., [17] proposed an effective technique to classify regions of interests (ROIs) of digitized mammograms into mass and normal tissue regions by first finding the significant texture features of ROI using binary PSO

(BPSO). The data set used consisted of sixty-nine ROIs from the MIAS Mini-Mammographic database. Eighteen texture features were derived from the GLCM of each ROI. Significant features are found by a feature selection technique based on BPSO. Experimental results showed that the significant texture features found by the BPSO based feature selection technique can have better classification accuracy when compared to the full set of features. The BPSO feature selection technique also has similar or better performance in classification accuracy when compared to other widely used existing techniques.

MATERIALS AND METHOD

This section discuss about Pseudo Zernike Moments and Gaussian Markov Random Field (GMRF) which are used for feature extraction. Hybrid Shuffled frog-PSO algorithm, IG for Feature selection and C4.5, Random Forest, Adaboost for Classifier.

PSEUDO ZERNIKE MOMENTS

The Zernike moments computation of an input image has 3 steps – computation of

a. radial polynomials,
b. Zernike basis functions and
c. Zernike moments by projecting image onto Zernike basis functions.

The kernel of pseudo-Zernike moments is orthogonal pseudo-Zernike polynomials set defined over polar coordinate space in a unit circle. The 2-dimensional pseudo-Zernike moments of order p with repetition q of an image intensity function is defined as:

$$Z_{pq} = \frac{p+1}{\pi}\int_{-\pi}^{\pi}\int_{0}^{1} V_{pq}^{*}(r,\theta)\, f(r,\theta)\, r dr d\theta;$$

$$|r| \le 1$$

where pseudo-Zernike polynomials V_{pq} of order p are defined as:

$$V_{pq}(r,\theta) = R_{pq}(r)e^{\hat{j}q\theta}; \quad \hat{j}=\sqrt{-1}$$

and the real-valued radial polynomials, $R_{pq}(r)$, is given as:

$$R_{pq}(r) = \sum_{k=0}^{p-|q|}(-1)^k \frac{(2p+1-k)!}{k!(p+|q|+1-k)!(p-|q|-k)!} r^{p-k}$$

Where $0 \le |q| \le p.$

As pseudo-Zernike moments are defined regarding polar coordinates (r,θ) with $|r| \le 1$, computation of pseudo-Zernike polynomials requires a linear transformation of image coordinates (i, j), i, j = 0, 1, 2, …, N-1 to a suitable domain $(x, y) \in R^2$ inside a unit circle. Two commonly used cases of transformations. Based on these, following discrete approximation of continuous pseudo-Zernike moments' integral [18].

$$Z_{pq} = \lambda(p,N)\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} R_{pq}(r_{ij})e^{-\hat{j}q\theta_{ij}} f(i,j),$$

$$0 \le r_{ij} \le 1$$

where most general image coordinate transformation to interior of unit circle is given by;

$$r_{ij} = \sqrt{(c_1 i + c_2)^2 + (c_1 j + c_2)^2}, \quad \theta_{ij} = \tan^{-1}\left(\frac{c_1 j + c_2}{c_1 i + c_2}\right)$$

GAUSSIAN MARKOV RANDOM FIELD (GMRF)

Let $x = (x_1, x_2, ..., x_n)^T$ be a Gaussian random field with mean μ and covariance matrix Σ, that is, $x \sim N(\mu, \Sigma)$. The precision matrix of x is denoted by Q and $Q = \Sigma^{-1}$. Gaussian random field x is said to be a Gaussian Markov Random Field (GMRF) regarding labeled undirected graph $G = (V, \varepsilon)$, if nodes are V = {1, …, n} and edges;

THE IIOAB JOURNAL ISSN: 0976-3104

$$\varepsilon = \{\{i, j\} \in V \times V : Q_{ij} \neq 0 \text{ and } \mathrm{i} \neq \mathrm{j}\}.$$

If $\{i, j\} \in \varepsilon$, then i and j are said to be neighbors and is written as $i \sim j$. Further, notation x_{ij} to refer to sub-vector of x corresponding to nodes i, i+1,…, j. By definition, any GRF is a GMRF, generally regarding a fully connected graph G.

In practice, use of GMRFs is confined to situations where neighborhood size is small so that precision matrix is sparse. The precision matrix's non-zero pattern is related to conditional independence structure of GMRF by $x_i \perp x_j \mid x_{-ij} \Leftrightarrow Q_{ij} = 0, i \neq j.$

Here, x_{-ij} denotes all elements of x except elements i and j. As a consequence of correspondence between non-zero pattern of Q and conditional independence structure of GMRF, GMRF is specified regarding its conditional moments.

A mammographic image Y is modeled by a finite lattice GMRF. Each pixel in image lattice L is represented by a random variable y_{ij} where $Y = \{y_{ij} : 0 \leq i \leq M-1, 0 \leq i \leq M-1\}$ and $L = \{(i, j) : 0 \leq i \leq M-1, 0 \leq i \leq M-1\}$. In a GMRF assumption of image Y with respect to a certain neighborhood system η, Y is reshaped to a single vector $y = [y_1, y_2, \ldots\ldots\ldots, y_M^2]$ in lexicographic order [18].

NFORMATION GAIN (IG)

Information Gain is supervised univariate feature selection algorithm of the filter model which is a measure of dependence between the feature and the class label. It is one of the most powerful feature selection techniques and it is easy to compute and simple to interpret. Information Gain (IG) of a feature X and the class labels Y is calculated as

$$IG(X, Y) = H(X) - H(X \mid Y)$$

Entropy (H) is a measure of the uncertainty associated with a random variable. H(X) and H(X/Y) is the entropy of X and the entropy of X after observing Y, respectively.

$$H(X) = -\sum_i P(x_i) \log_2 (P(x_i)).$$

The maximum value of information gain is 1. A feature with a high information gain is relevant. Information gain is evaluated independently for each feature and the features with the top-k values are selected as the relevant features. This feature selection algorithm does not eliminate redundant features [19].

$$H(X \mid Y) = -\sum_j P(y_j) \sum_i P(x_i \mid y_j) \log_2 (P(x_i \mid y_j))$$

PROPOSED SHUFFLED FROG ALGORITHM-PARTICLE SWARM OPTIMIZATION FOR FEATURE SELECTION

Particle swarm optimization algorithm is an optimization algorithm based on group and fitness. The system initializes particles (representing potential solutions) as a set of random solutions, which has two features of position and velocity. The fitness values of particles are decided by particle positions. Particles move in the solution space; the moving direction and distance are determined by the speed vector and new speed, position are updated from personal best position pbest, global best position gbest and the current particle velocity; particles search and pursue the optimal particle based on fitness values in the solution space, and gradually converge to the optimal solution. Assuming in a d-dimensional search space, there is a group composed of n particles, where of generation t particle i (i = 1, 2, ..., n), position coordinates $x_i^t = (x_{i1}, x_{i2}, \ldots, x_{id})$, velocity $v_i^t = (v_{i1}, \mathrm{v}_{i2}, \ldots, \mathrm{v}_{id})$ personal best position $p_i^t = (p_{i1}, \mathrm{p}_{i2}, \ldots, \mathrm{p}_{id})$ and global best position $p_g^t = (p_{g1}, \mathrm{p}_{g2}, \ldots, \mathrm{p}_{gd})$. For particle i dimension d generation t, its iterative formula can be expressed as:

$$v_{id}^{t+1} = \omega \vartheta_{id}^t + c_1 r_1 (p_{id}^t - x_{id}^t) + c_2 r_2 (p_{gd}^t - x_{id}^t)$$

$$x_{id}^{t+1} = x_{id}^t + \vartheta_{id}^{t+1}$$

where

ϑ_{id}^t - Current velocity,

ϑ_{id}^{t+1} - New speed of particle r after iteration t,

ω - Inertia weight,

c_1, c_2 - Acceleration (learning) factors,

r_1, r_2 - Uniformly distributed random numbers between 0 and 1,

x_{id}^{t} - Current position of particle i,

x_{id}^{t+1} -new position of particle i after iteration t.

Shuffled frog leaping algorithm is a biological evolution algorithm based on swarm intelligence. The algorithm simulates a group of frogs in the wetland passing thought and foraging by classification of ethnic groups. In the execution of the algorithm, F frogs are generated at first to form a group, for N-dimensional optimization problem, frog i of the group is represented as $\left(x_i^1, x_i^2, \ldots, x_i^N\right)$ then individual frogs in the group are sorted in descending order according to fitness values, to find the global best solution P_x . The group is divided into m ethnic groups, each ethnic group including n frogs, satisfying the relation F=m×n. The rule of ethnic group division is: the first frog into the first sub-group, the second frog into the second sub-group, frog m into subgroup m, frog m+1 into the first sub-group again, frog m+2 into the second sub-group, and so on, until all the frogs are divided, then find the best frog in each subgroup, denoted by P_b ; get a worst frog correspondingly, denoted by P_ω . Its iterative formula can be expressed as:

$$D = rand()^{*}\left(P_b - P_\omega\right)$$

$$P_{new-\omega} = P_\omega + D_i, -D_{\max} \le D_i \le D_{\max}$$

where $rand()$ represents a random number between 0 and 1,

P_b represents the position of the best frog,

P_ω represents the position of the worst frog,

D represents the distance moved by the worst frog,

$P_{new-\omega}$ is the improved position of the frog,

Dmax represents the step length of frog leaping.

In the execution of the algorithm, if the updated $P_{new-\omega}$ is in the feasible solution space, calculate the corresponding fitness value of Pnew-w, if the corresponding fitness value of $P_{new-\omega}$ is worse than the corresponding fitness value of P_ω , then use P_ω to replace P_b and re-update $P_{new-\omega}$; if there is still no improvement, then randomly generate a new frog to replace P_ω ; repeat the update process until satisfying stop conditions.

Exploration and exploitation has been a contradiction in the search process of swarm intelligence algorithms. Exploration stresses searching for a new search region in the global range, and exploitation is focused on fine search in local search area. Although particle swarm optimization algorithm is simple and its optimization performance is good, in the entire iterative process, exploration capability is strong and exploitation capability is weak in early period, at this time if particles fall on the neighbourhood of the best particle, they may flee the neighbourhood of the best particle, due to too strong exploration capability; exploration capability is weak and exploitation capability is strong in later period, at this time if particles encounter local optima, the speed of all particles may be rapidly reduced to zero instead of flying, leading to convergence of particle swarm to local optima; the iterative mechanism and ethnic group division lead to strong exploitation and weak exploration in early period, and strong exploration and weak exploitation in later period.

Based on the analysis, in the update process of the algorithm, in order to ensure the diversity of particles, particle swarm and frog group sharing part of the particles, we propose particle sharing based particle swarm frog leaping hybrid optimization algorithm. The idea is as follows: divide the total number of particles N into two sub-groups of numbers N1 and N2, where the first sub-group uses shuffled frog leaping algorithm to optimize, the second sub-group uses the standard particle swarm optimization algorithm to optimize, and N, N1 and N2 satisfy N≤N1+N2, so the number of shared particles is N1+N2－N [20].

CLASSIFIER

Classification models are monitored methods that are initially trained on a dataset of samples known as training sets. The performance of the algorithms is then evaluated on distinct training sets. The features that are extracted are inputs for the classifiers. The performance of three classifiers is examined on datasets.

C4.5

C4.5 is an extension of Iterative Dichotomizer (ID3) algorithm that was designed by Quinlan to deal with issues that cannot be handled by the ID3 algorithm. These include avoidance of over fitting the data; reduced error pruning, rule post-pruning, handling continuous attributes and handling data with missing attribute values. It attempts to build a decision tree with a measure of the

ISSN: 0976-3104

information gain ratio of each feature and branching on the attribute which returns the maximum information gain ratio. Pruning takes place in C4.5 by replacing the internal node with a leaf node thereby reducing the error rate [21].

Typically, C4.5 assigns the frequency of the correct counts at the leaf as the probabilistic estimate. For notational purposes, TP is the number of true positives at the leaf, FP is the number of false positives, and C is the number of classes in the data set. Thus, the frequency based probabilistic estimate can be written as [22]:

$$P_{leaf} = TP/(TP+FP)$$

RANDOM FOREST

Random Forest (RF) is an approach which has been proposed by Breiman for classification tasks. It mainly comes from the combination of tree-structured classifiers with the randomness and robustness provided by bagging and random feature selection. The classification is performed by sending a sample down is each tree and assigning it the label of the terminal node it ends up in. At the end the average vote of all trees is reported as the result of the classification. Random forest is very efficient with large datasets and high dimensional data [21].

The principle of RF is the aggregation of a large ensemble of decision trees. During training, each individual tree in the ensemble is fitted by sampling the training data with replacement (bootstrap) and growing the tree to full depth on the training sample. The optimal data split at each tree node is determined by randomly choosing m of the available P input variables and selecting the one which splits the node best. In this work, node splitting was guided by the Gini cost function

$$G(N) = 1 - \sum_{k=1}^{2} p^2(\omega_i)$$

which measures node impurity using $p(\omega_i)$ as the fraction of features in class i at node N. The best split was the one which decreases node impurity the most. Further, by calculation of the mean decrease in Gini (MDG) for each variable over all trees, RF allow to obtain a variable importance ranking. The final RF classification score is determined by collecting the votes of each of the n trees in the forest for either class and outputting a vote ratio. As the method is based on decision trees, the splits in the nodes are always parallel to the coordinate axes of the features [23].

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. Also it is used for maintaining a set of weights over the training set. The training set $(x_1, y_1), \ldots\ldots.(x_n, y_n)$ where each xi belongs to instance space X and each yi is in the label set Y= {-1, +1}. The steps for AdaBoost are as follows [24]:

1. Assign N example

   $(x_1, y_1), \ldots..(x_n, y_n); x_i \in \{-1, +1\}$
2. Initialize the weights of $D_1(i) = 1/N, i = 1, \ldots..N$
3. For $k = 1, \ldots K$
4. Train weak learner using distribution D_k
5. Get weak hypothesis $h_k : X \rightarrow R$ with its error: $\varepsilon_k = \sum_{i=h_k(x_i)\neq y_i} D_k(i)$
6. Choose $\varepsilon_k = R$
7. Update

$$D_{k+1}(i) = \frac{D_k(i)\exp(-\alpha_k y_k k_k(x_k))}{z_k}$$

where z_k is the normalization factor.

8. Output the final hypothesis:

$$H(x) = sign\left(\sum_{k=1}^{K} \alpha_k h_k(x)\right).$$

RESULTS

ISSN: 0976-3104

To evaluate the proposed technique, 150 normal mammogram image and 25 images with calcification obtained from MIAS dataset were used. Features are extracted using Pseudo Zernike Moments and Gaussian Markov Random Field technique. Features are selected using IG and the proposed hybrid SF-PSO. Classification is achieved using C4.5, random tree and AdaBoost techniques. Results are presented in this section. [Table- 1] and [Figure- 1], shows the classification accuracy.

Table: 1. Classification Accuracy

Techniques	IG	Hybrid SF-PSO
C4.5	84	94.97
Random tree	84.57	95.53
Boosting	85.14	97.21

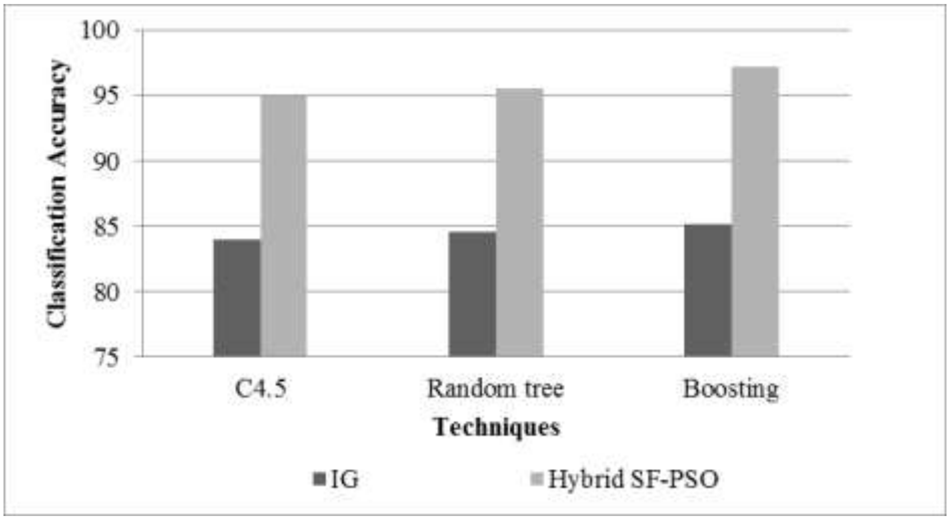

Fig: 1. Classification Accuracy

From [Table- 1] and [Figure- 1], it can be observed that the classification accuracy has improved for hybrid SF-PSO than IG by an average of 12.56%. For C4.5, hybrid SF-PSO has improved classification accuracy by 12.26% than IG. Similarly for Random tree, hybrid SF-PSO has improved classification accuracy by 12.17% than IG and for Boosting, hybrid SF-PSO has improved classification accuracy by 13.24% than IG. [Table- 2] and [Figure- 2] shows the sensitivity.

Table: 2. Sensitivity

Techniques	IG	Hybrid SF-PSO
C4.5	0.76	0.88
Random tree	0.8	0.92
Boosting	0.8	0.92

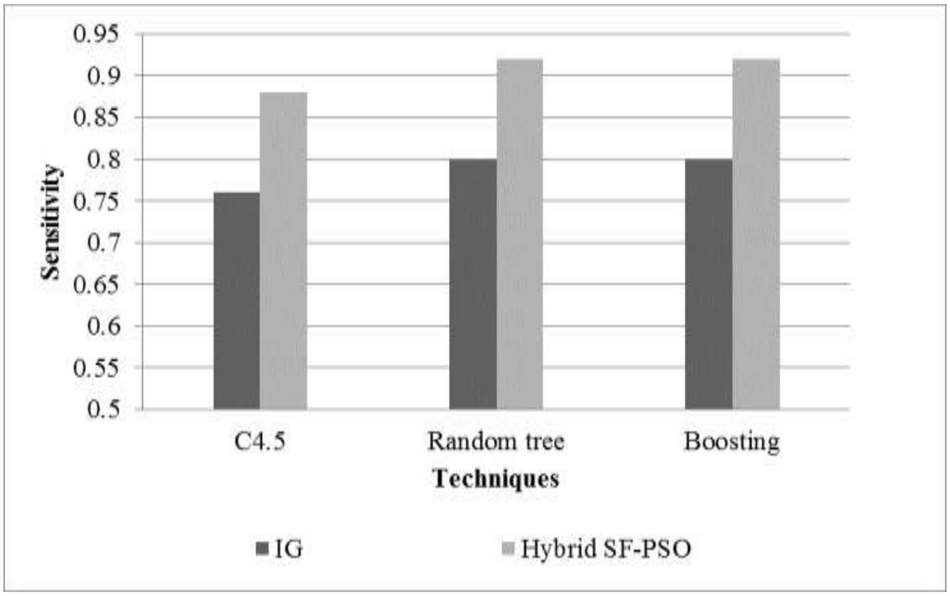

Fig: 2. Sensitivity

From [Table- 2] and [Figure- 2], it can be observed that the sensitivity has improved for hybrid SF-PSO than IG by an average of 14.17%. For C4.5, hybrid SF-PSO has improved sensitivity by 14.63% than IG. Similarly for Random tree, hybrid SF-PSO has improved sensitivity by 13.95% than IG and for Boosting, hybrid SF-PSO has improved sensitivity by 13.95% than IG. [Table- 3] and [Figure- 3] shows the specificity.

Table: 3. Specificity

Techniques	IG	Hybrid SF-PSO
C4.5	0.8533	0.961
Random tree	0.8533	0.961
Boosting	0.86	0.9805

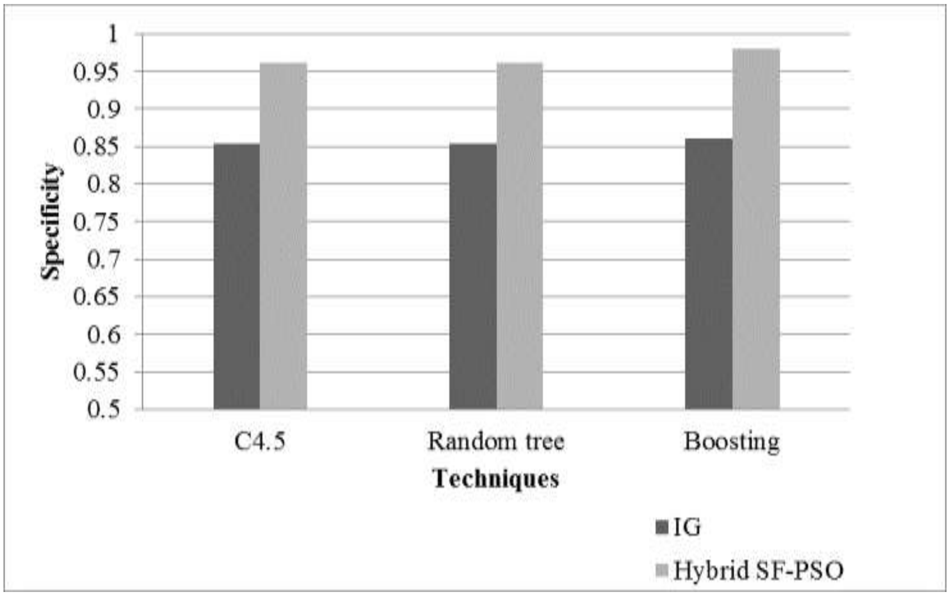

Fig: 3. Specificity

From [Table- 3] and [Figure- 3], it can be observed that the specificity has improved for hybrid SF-PSO than IG by an average of 12.29%. For C4.5 and Random tree, hybrid SF-PSO has improved specificity by 11.87% than IG. Similarly for Boosting, hybrid SF-PSO has improved specificity by 13.09% than IG. [Figure- 4] shows the Percentage of features selected – Hybrid SF-PSO.

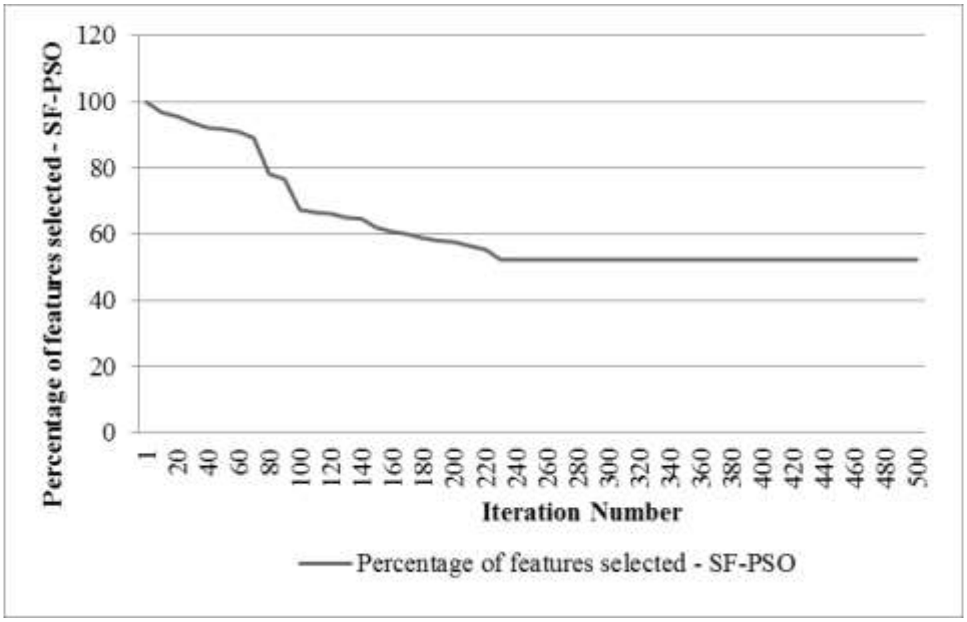

Fig: 4. Percentage of features selected – Hybrid SF-PSO

………

In iteration 230, 52% of features are selected which forms the optimal feature subset.

CONCLUSION

Breast cancer is one of the most common cancers among women around the world. Mammography is one of the best breast cancer detection methods. But, in some cases, radiologists face problems in detecting tumours. A feature extraction method for finding the most significant coefficients was proposed and implemented to classify a set of mammogram images. This study presented a new approach to segment breast cancer mass in mammograms. The study focuses on improving classification performance through feature selection. It is seen that of the various classification techniques C4.5 outperforms other algorithms with highest accuracy. Using AdaBoost followed by domain adjusted post-processing such as false positive filtering, our approach achieved promising preliminary results. The classification accuracy has improved for hybrid SF-PSO than IG by an average of 12.56%. For C4.5, hybrid SF-PSO has improved classification accuracy by 12.26% than IG. Similarly for Random tree, hybrid SF-PSO has improved classification accuracy by 12.17% than IG and for Boosting, hybrid SF-PSO has improved classification accuracy by 13.24% than IG.

.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ACKNOWLEDGEMENT

None

FINANCIAL DISCLOSURE

The authors report no financial interests or potential conflicts of interest.

REFERENCES

[1] Nithya R, Santhi B. [2011] Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer. *International Journal of Computer Applications*, *28*(6): 21-25.

[2] Bosch A, Munoz X, Oliver A, Marti J. [2006] Modeling and classifying breast tissue density in mammograms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* 2: 1552-1558). *IEEE*.

[3] Sampat M.P, Markey MK, Bovik AC. [2005] Computer-aided detection and diagnosis in mammography. *Handbook of image and video processing 2*(1):1195-1217.

[4] Nithya R, Santhi B. [2011] Comparative study on feature extraction method for breast cancer classification. *Journal of Theoretical and Applied Information Technology*, *33*(2):1992-1986.

[5] Ramani R, Vanitha NS. [2014] Computer A ided Detection of Tumours in Mammograms. *International Journal of Image, Graphics and Signal Processing 6*(4): 54.

[6] Arpana MA, Kiran P. [2014] Feature Extraction Values for Digital Mammograms. *International Journal of Soft Computing and Engineering (IJSCE) 4*(2): 183-187.

[7] Luo ST, Cheng BW. [2012] Diagnosing breast masses in digital mammography using feature selection and ensemble methods. *Journal of medical systems*, *36*(2):569-577.

[8] Eltoukhy MM, Faye I. [2014] An Optimized Feature Selection Method For Breast Cancer Diagnosis in Digital Mammogram using Multiresolution Representation. *Appl. Math 8*(6): 2921-2928.

[9] Rehman AU, Chouhan N, Khan A. [2015] Diverse and Discrimintative Features based Breast Cancer Detection using Digital Mammography. In *2015 13th International Conference on Frontiers of Information Technology (FIT)* (pp. 234-239). IEEE.

[10] Patel BC, Sinha GR. [2014] Mammography feature analysis and mass detection in breast cancer images. In *Electronic Systems, Signal Processing and Computing Technologies (ICESC), 2014 International Conference on* (pp. 474-478). IEEE.

[11] Ganesan K, Acharya UR, Chua CK, Lim CM, Abraham KT. [2014] One-class classification of mammograms using trace transform functionals. *IEEE Transactions on Instrumentation and Measurement 63*(2): 304-311.

[12] Deshpande DS, Rajurkar AM, Manthalkar RM. [2013] Medical image analysis an attempt for mammogram classification using texture based association rule mining. In *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on* (pp. 1-5). IEEE.

[13] Sanae B, Mounir AK, Youssef F. [2014] Statistical block-based DWT features for digital mammograms classification. In *Intelligent Systems: Theories and Applications (SITA-14), 2014 9th International Conference on*(pp. 1-7). IEEE.

[14] Kim S. [2014] Margin-maximised redundancy-minimised SVM-RFE for diagnostic classification of mammograms. *International journal of data mining and bioinformatics*, *10*(4), 374-390.

[15] Thangavel K, Velayutham C. [2012] Rough set based unsupervised feature selection in digital mammogram image using entropy measure. In *Biomedical Engineering (ICoBE), 2012 International Conference on* (pp. 10-16). IEEE.

[16] Aroquiaraj IL, Thangavel K. [2013] Mammogram image feature selection using unsupervised tolerance rough set relative reduct algorithm. In *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on* (pp. 479-484). IEEE.

[17] Wong MT, He X, Nguyen H, Yeh WC. [2012] Particle swarm optimization based feature selection in mammogram mass classification. In *Computerized Healthcare (ICCH), 2012 International Conference on* (pp. 152-157). IEEE.

[18] Devisuganya, S, & Suganthe, R. C. [2016] Breast Cancer Detection: A Framework to Classify Mammograms.

[19] Porkodi R. [2014] comparison of filter based feature selection algorithms: An overview. *international journal of innovative research in technology& science*, *2*(2): 108-113.

[20] Lenin K, dranath Reddy BR, Kalavathi MS. [2014] Particle Sharing Based Particle Swarm Frog Leaping Hybrid Optimization Algorithm for Solving Optimal Reactive Power Dispatch Problem.

[21] Oleiwi ASA. [2014] Classification of Mammography Image Using Machine Learning Classifiers and Texture Features.

[22] Chawla NV. [2003, August] C4. 5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure. In *Proceedings of the ICML* (Vol. 3).

[23] Lesniak JM, Hupse R, Blanc R, Karssemeijer, N & SzékelyG. [2012] Comparative evaluation of support vector machine classification for computer aided detection of breast masses in mammography. *Physics in medicine and biology*, *57*(16): 5295.

[24] Ramani R, Vanitha NS. [2015] Computer Aided Detection Of Tumors in Mammograms using Optimized Support Vector Machines. ARPN *Journal of Engineering and Applied Sciences*, 10(4).