

An enhancement of mammogram images for breast cancer classification using artificial neural networks

Jalpa J. Patel¹, S. K. Hadia²

¹V.T. Patel Department of Electronics and Communication Engineering, Chandubhai S. Patel Institute of Technology, Charotar University of Science and Technology, Changa, Gujarat, India.

²Gujarat Technological University, Ahmedabad, Gujarat, India

Article Info

Article history:

Received Oct 22, 2020

Revised Feb 27, 2021

Accepted Apr 1, 2021

Keywords:

CLAHE

FFBPNN

GLCM

HOFS

Morphological operations

Multithresholding

ABSTRACT

Breast cancer is the most driving reason for death in women in both developed and developing nations. For the plan of effective classification of a system, the selection of features method must be used to decrease irregularity part in mammogram images. The proposed approach is used to crop the region of interests (ROIs) manually. Based on that number of features are extracted. In this proposed method a novel hybrid optimum feature selection (HOFS) method is used to find out the significant features to reach maximum accuracy for this classification. A number of selected features is applied to train the neural network. In this proposed method accessible informational index from the mini-mammographic image analysis society (MIAS) database was used. The classification of this mammogram database involved a neural networks classifier which attained an accuracy of 99.7% with a sensitivity of 99.5%, and specificity of 100% as the area under the curve (AUC) is 0.9975 and matthew's correlation coefficient (MCC) represents a binary class value which reached the value of 0.9931. It can be useful in a computer-aided diagnosis system (CAD) framework to help the radiologist in analyzing breast cancer. Results achieved with the proposed method are better compared to recent work.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Jalpa J. Patel

V.T. Patel Department of Electronics and Communication Engineering

Chandubhai S. Patel Institute of Technology

Charotar University of Science and Technology, Changa, Gujarat, India

Email: jalpapatel.ec@charusat.ac.in

Abbreviations: ROIs: region of interests, MIAS: mammographic image analysis society, CAD: computer aided diagnosis, DDSM: digital database for screening mammography, CT: contourlet, EML: extreme machine learning technique, GLCM: gray level co-occurrence matrix, GLRLM: gray level run length matrix, PSO-SVM: particle swarm optimization-support vector machine, ANFIS: adaptive neuro fuzzy interface system, AHE: adaptive histogram equalization, CLAHE: contrast-limited adaptive histogram equalization, ANNs: artificial neural networks, HOFS: hybrid optimum feature selection, FFBPNN: feed forward with back propagation neural network, SVM: support vector machine, FC: fuzzy classification, AUC: area under the curve, ROC: receiver operating curve, FFNN: feed forward neural network, Ls-SVM: linear support vector machine, MLP: multi-layer perceptron, PSOWNN: particle swarm optimization wavelet neural network, RS: region segmentation, TPR: true positive rate, FPR: false positive rate, MCC: matthews correlation coefficient.

1. INTRODUCTION

Breast cancer is a common disease affecting women all over the world due to its demise rate. As estimated by the India cancer society based on their globocon data 2018 [1], the new cases registered of breast cancer are 1, 62,468 among women and 87,090 deaths. It is the second-highest in the world for the year 2018. The subsistence rate of breast cancer in India is very small because the recognition happens very late. As far as mortality, it is the main cause of death among ladies matured somewhere in the range of 35 to 64 years and is the main disease-related for death in the female population [2], [3]. According to globocan data, Figure 1 shows the statistics of all cancers for all ages women. It is the most ordinary cancer in the world and there is a consistent increment in breast cancer growth cases among young ladies [4].

Hence, the number of deaths can be reduced by detecting breast cancer in its early stages. Numerous screening methods for breast cancer are existing like mammography, Positron emission tomography, magnetic resonance imaging, and ultrasonography. Out of that mammography is considered the most trustworthy and it is also an efficient method for the initial stage detection of breast cancer. Mammography used for breast cancer can be characterized into two classes-normal and abnormal, and the irregular mass category can be separated into two classes-non-carcinogenic and malignant. Benevolent mass isn't damaging to wellbeing; their cells have a nearby preference to typical prosenchyma. Benign mass develops moderately gradually and doesn't attack the contiguous tissues that spread to various parts of the body [5]-[7]. Identification of some infection depends on the human experience, the significant as several of the medical diseases. It takes a long time to process and has a human error in the results. Some works are based on computational intelligence approaches such as, ANNs which are used in several areas as a great computational intelligent technique is presented in [8]-[9].

Even though ANNs have been very fruitful, there is a necessity to improve and optimize the ANNs in terms of overall results and accuracy. Computerized classifiers may be suitable for radiologists in differentiating between normal, non-carcinogenic, and carcinogenic patterns. Thus, in this research paper, ANNs which can be helped as a computerized classifier are examined. In the field of medical image processing, ANNs have been useful to a variety of data classification and pattern recognition and have become a favorable classification tool in breast cancer [10]. For cancer identification, segmentation, and classification of an image ANNs have been widely used. Several types of image segmentation methods, based on histogram features, texture features, geometric features, and statistical features, have been trained using ANNs [11]. Image features can be extracted in many parts, like statistical, texture, and shape. Thus, different types of selected image features will give different classification results.

In this research paper, an inventive and genuine classification methodology based on the ROIs segmented lesion is proposed. Initially, breast lesion is implemented for segmentation using the multi level thresholding method. Furthermore, extracting features such as statistical, shape, texture, and geometric from the segmented suspicious region and its backgrounds and then to set up an optimum feature selected based on the extracted features, finally they are differentiated as a normal and abnormal mass based on the recognized feature database with the classifier of FFBPNN.

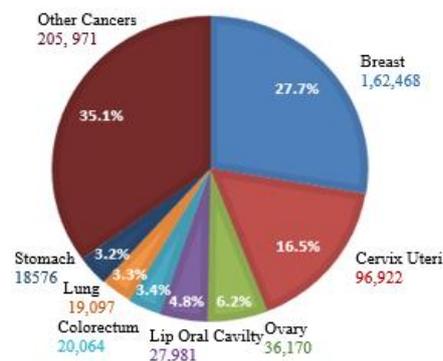


Figure 1. Estimated number of new cases in 2018, for all ages of females in India [1]

2. RELATED WORK

Shankar Thawker *et al.* [12] projected a novel ANFIS for classification. To enhance efficient feature selection method using bio-geography based optimization selection for the subclass of feature selection. The proposed method was investigated using the DDSM database for better classification accuracy, sensitivity, specificity, and AUC. Nadeem Tariq *et al.* [13] described the identification of benign and malignant mass for

designing CAD systems. In the proposed system, they have found out texture features from mammogram images to calculate GLCM with 0° , to calculate several features for desired output, and applied to ANN for training the networks for this binary classification. The proposed system benefited the radiologist to increase the identification of tumor accuracy. The system achieved an accuracy of 99.4% with a sensitivity of 100% and a specificity of 99.4% for the MIAS database.

Amit Kamra *et al.* [14] proposed various types of ROIs selected for diagnosis of breast cancer for diagnosis using shape and texture-based techniques. The number of ROIs and their location proposed for CAD systems. In his research work, according to several ROIs size to categorized dense and fatty mammogram images from MIAS database. Using, fisher discriminant ratio for selection feature for classification. Finally, linear SVM was applied for the classification which achieved the accuracy of 96.1% with ROIs size 200*200 pixels.

Weiyang Xie *et al.* [15] formulated EML for grouping of benign and malignant mass. The multidimensional feature vector was constructed. For feature, the selection was made using SVM and EML. Lastly, the optimum subset of feature vectors is input to the classifier and find out benign mass and malignant mass. They additionally contrast their proposed framework with PSO-SVM and achieved efficient execution in the proposed CAD system. The proposed system attained the accuracy of 95.73% and 96.02% on the MIAS and DDSM database, respectively. Anuj Kumar Singh *et al.* [16] proposed maximum-mean and least-variance technique for mass identification for breast cancer diagnosis.

S. Punitha *et al.* [17] described an automatic recognition for breast masses to utilize the underlying seed focus on region growing technique to accurately diagnosing using optimum thresholding generated by swarm optimization techniques called as dragonfly algorithm. For feature extraction, GLCM and GLRLM methods are used for texture-based features. A number of selected features are fed to the FFNN algorithm is applied to identify cancerous and non-cancerous mass. For this proposed method sensitivity and specificity achieve 98.1% and 97.8% respectively for the DDSM database. Rahimeh Rouhi *et al.* [18] has proposed automatic mass detection for primary detection of breast cancer using region growing methods, for training this algorithm cellular neural network with the genetic algorithm used. And achieved 96.47% accuracy with 96.87% sensitivity and 95.94% specificity for MIAS and DDSM database.

Arnau Oliver *et al.* proposed various texture feature selected utilizing the GLCM algorithm and applied on sub-images to improve its performance. Here they can discover out statistical features and according to numbers of feature vector proposed wavelet-CT1, wavelet-CT2 and ST-GLCM also compare to multi-resolution features [19]. Another method uses Principal Component Analysis for classification of ROIs images. They tested performance on two subdivisions of MIAS database. They achieved area under the curve values is 0.74 and accuracy is 77% [20]. The significant issue is the minor number of tests around 20 abnormal and other 20 for ordinary images are utilized for assessment which is not adequate to the analysis of their methods. GLCM is additionally utilized for breast lesions characterization at four directions of $\theta (\pi/4, \pi/2, 3\pi/4)$ [21]. A Mohd. Khuzi *et al.* projected a novel method that uses principal component analysis for classification of ROIs images. They tested performance on two subdivisions of MIAS and DDSM databases. The detailed area under the curve values is 0.84 for the MIAS dataset [22]. Table 1 shows the highlights of the features considered in the current strategies, alongside the difficulties that should have been defeated in the future, for the correct expectation of breast cancer.

Table 1. Feature and tasks of predictable breast cancer analysis models

Authors	Methodology used	Features	Challenges
Shankar Thawker <i>et al.</i> [12]	ANFIS	<ul style="list-style-type: none"> ▪ Computation time is high ▪ Highest classification accuracy 	<ul style="list-style-type: none"> ▪ Parameters selected based on trial-and-error method
Nadeem Tariq <i>et al.</i> [13]	ANN	<ul style="list-style-type: none"> ▪ Number of Texture and statistical features are extracted. ▪ To diagnosis high accuracy 	<ul style="list-style-type: none"> ▪ Regression analysis is done by the number of extracted features.
Amit Kamra <i>et al.</i> [14]	Linear SVM	<ul style="list-style-type: none"> ▪ Texture Feature Extracted ▪ Accuracy is high 	<ul style="list-style-type: none"> ▪ Enhance resolution towards the nipple area ▪ To reduce computation time
Weiyang Xie <i>et al.</i> [15]	EML	<ul style="list-style-type: none"> ▪ Highest area under curve ▪ Achieve highest accuracy 	<ul style="list-style-type: none"> ▪ Lesion segmentation by level set function ▪ Applicable for cancerous and non-cancerous mass
Anuj kumar <i>et al.</i> [16]	Maximum-Mean and Least-Variance technique	<ul style="list-style-type: none"> ▪ Identify tumor region from mammogram images ▪ Averaging and Thresholding 	<ul style="list-style-type: none"> ▪ Manual selection of threshold parameter ▪ Lack of transparency of results
S. Punitha <i>et al.</i> [17]	FFNN	<ul style="list-style-type: none"> ▪ To diagnose high accuracy ▪ Guide to radiologists for early detection of tumor 	<ul style="list-style-type: none"> ▪ Optimize region growing algorithm ▪ Applicable for malignant and benign lesion

3. DATA SET AND PROPOSED METHOD

3.1. Data set

All the mammogram images utilized in this paper are having a freely available data set of mammographic images. The first dataset is looked over the Mini MIAS database in the United Kingdom [23]. Additionally, the pictures were in the PGM record group and 1024*1024 pixels with an 8-bit gray level. MIAS contains 322 mammograms which are categorized into normal, benign, and malignant classes. Of the database, 208 mammograms are regular, while the other 115 signify abnormal. There are 63 irregular mammograms that are benign, and the other 52 are malignant. Image number mb258, mb260, and mb297 are not considered in our proposed technique because of the low quality of the images.

3.2. Proposed method

A CAD methodology for the ROIs based breast cancer detection is proposed in this paper. The proposed method helps to the analytic process of breast cancer classification. Figure 2 shows a diagram representation of the diagnosis of breast cancer detection of the proposed method.

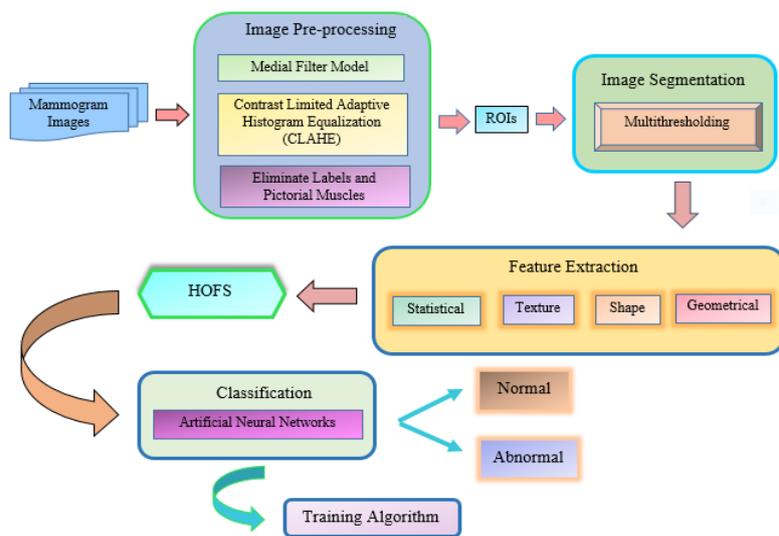


Figure 2. Diagram representation of proposed diagnosis of breast cancer model

3.2.1. Image pre-processing

It is the primary stage, and it executes adaptive median filtering and CLAHE to the input mammogram images.

a. Median filter model

Initially, it removes unwanted parts or irrelevant areas from the background of the mammogram images. It also improves the quality of mammogram images. The median filter is used for the pre-processed image. It is a nonlinear filter utilized in a specific area of an image. It investigates the image pixel by pixel and replaces every pixel with the middle of neighbouring entries. It is also used to remove artifacts and background to remove noise. Then apply an adaptive waited median filter was applied to the median filter to improve the quality of the input image.

b. CLAHE

For image improvement, AHE is utilized. It improves the contrast of the grayscale image by changing the characteristics using CLAHE. AHE adjusts image intensity in an extremely little area of the image [24]-[25].

c. Eliminate label and pictorial muscles

To eliminate the extra part of the image, in this proposed work binary upper triangle method is used to remove labels and the redundant parts of mammogram images.

3.2.2. Image segmentation

Doubtful lesions are identified using the segmentation method from the mammogram images for more processing. The mass tends to be brighter than the neighbouring area, therefore they have higher intensity values. Multithresholding method based on Otsu's method segments a gray level image in excess of a couple of independent areas. It is a method that segments a gray level image in excess of a couple of independent areas. In this method, we decide more than one threshold for the given image and segments the image into a certain bright area, which is related to one background and more than a few substances. For that morphological functions, opening and closing are used [26], [27].

3.2.3. Feature extraction

It plays a powerful job in pattern classification.

a. The first order statistical feature method

This procedure is utilized for finding the intensity level of the histogram for mammogram images to determine the gray level intensity parameters. Six statistical features are extracted from the ROIs- mean, variance, skewness, kurtosis, energy, and entropy [28]. The dissimilarity of ROIs is calculated using these statistical features according to their brightness and contrast.

b. Shape-based method

The shape and edge of the mass are in categorizing images as normal, benign and malignant. Shape-based features are area, circularity, perimeter, compactness, uniformity, roundness, and solidity have been calculated from the selected ROIs [28]. These types of features are used to determine the circularity of the mass and their abnormality.

c. Texture-based feature method

It is one of the procedures to obtain the second-order statistical features. GLCM is a popular texture-based feature extraction method. Using the GLCM method determine the texture relationship between pixels by performing an operation according to the second-order statistics in the images. We can also determine correlation, homogeneity, contrast and energy for each pixel of second-order texture features from the normalized GLCM of the image. 16 features are extracted using GLCM method along with θ is 0° , 45° , 90° , and 135° with distance $d=1$ used for in pixels and orientation [29]. We have also measured texture features like texture mean, texture global mean, texture Standard deviation, texture smoothness, texture entropy, texture skewness, and texture correlation [29].

d. Geometric based feature method

Geometric features are seen as viable in segregating normal, benign and malignant masses. Similarly, shape-based features can be classified based on the shape of the mass. It has been seen that the shape of a mass is viewed as moderately not clear. This is because of the way that normal and abnormal class of masses emerge from one spot and develop peripheral. The abnormal mass has round and circular shapes since they suggest a well circumferentially. Figure 3 defined a segmented region for this proposed method and separate geometric shape descriptors such a Ge. Area, Ge. Perimeter and Ge. Compactness [30]. We can also find out another element global mean.

All these features are very important for differentiating normal and abnormal lesions. The results of enhanced mammogram images using this proposed method shown in Figure 3. Here, Figure 3(a) shows an original mammogram (MB184) which is containing benign lesions with image enhancement shown in Figure 3(b) (enhanced image-1), it is observed that the background region is suppressed with contrast is improved. Then ROI as shown in Figure 3(c) (ROI-1), shows that targeted ROI is very clearly identified. In Figure 3(d) (segmented image-1); here, shows segmented parts of our original image. Similarly, Figure 3(e) (MB005) shows the original and Figure 3(f) (enhanced image-2) enhanced mammogram images with ROI shows in Figure 3(g) (ROI-2) and the segmented image shown in Figure 3(h) (segmented image-2).

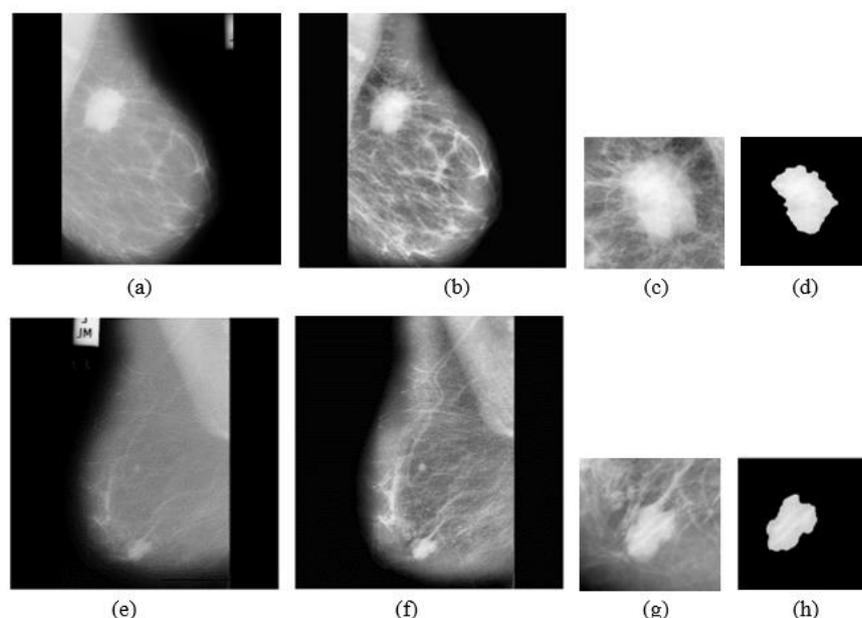


Figure 3. Original image-1 Mb184 and enhanced mammograms with subsequent ROI and segmented image part. (a) Mam-1, (b) Enhanced image-1, (c) ROI-1, (d) Segmented image-1, (e) Original image-2 Mb005, (f) Enhanced image-2, (g) ROI-2, (h) Segmented image-2

3.2.4. Proposed hybrid optimum feature selection method

In this proposed HOFS method is used to select the optimum features set. Proposed solution using HOFS, best feature vector F is selected. In this algorithm for input dataset, processed with assumed higher redundancy, can be converted into less features set. Based on the extracted feature, information gain can be calculated by comparing the entropy of the dataset before and after a transformation and check their rank. After this process, high-rank features are identified according to the rank feature method. The main objective of these proposed method is to diagnostic maximum accuracy. For each feature, properties are extracted and developed as feature vector is given underneath:

$$F = \{F1, F2, F3, F4\} \quad (1)$$

Where F dimension is 40 extracted features from masses of mammogram images. Figure 4 represents the proposed method in which the novel HOFS used with the combination of information gain and rank feature selection method provides better targeted output. It has been found that using less number of feature accuracy is high compared to the high number of features. In this paper, HOFS provides the most valuable 38 feature vectors which are chosen for classification. This part makes centre around acquiring a subset of F that can accomplish better execution of classification [31], [32]. In this paper, mammogram images are labeled as normal or abnormal lesion. Although each lesion type has its specific characteristics, it is very difficult to separate between normal and abnormal mass. In general, cells in cancerous tissues tend to expand and becomes closer. In normal mass, the cell forms are more regular and the color is black. So, the classification of complex mammogram images, this type of features has been selected. Table 2 represents the most significant number feature for the greatest region under the curve for the higher characterization of exactness.

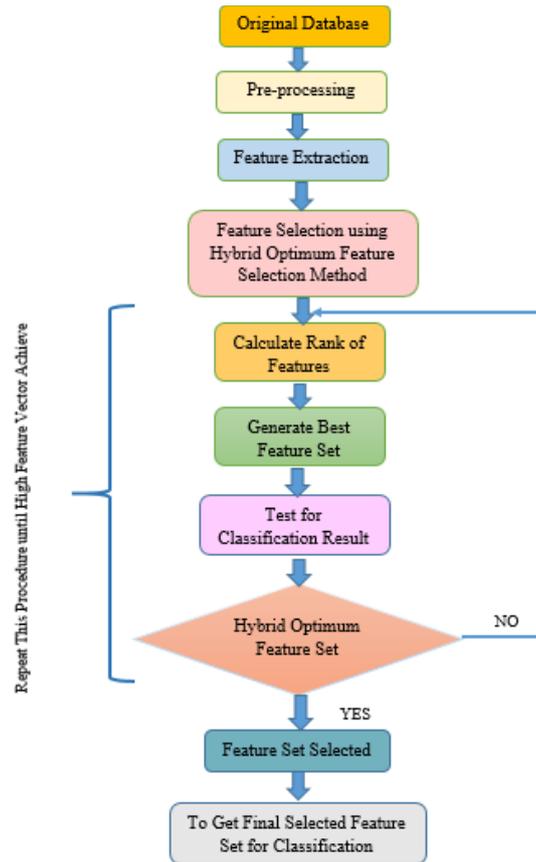


Figure 4. Hybrid optimum feature selection model

Table 2. Selected feature vector number for individual feature

Feature vector	Feature	Number	Feature vector	Feature	Number
F ₁ ¹	Mean (Mass)	0.7845	F ₃ ⁸	Energy GLCM at 0°	0.5154
F ₁ ²	Standard Deviation (Mass)	0.9580	F ₃ ⁹	Contrast GLCM at 0°	0.1605
F ₁ ³	Variance (Mass)	0.9458	F ₃ ¹⁰	Correlation GLCM at 0°	0.9931
F ₁ ⁴	Kurtosis	0.0259	F ₃ ¹¹	Contrast GLCM at 45°	0.3463
F ₁ ⁵	Smoothness	0.9997	F ₃ ¹²	Correlation GLCM at 45°	0.9853
F ₁ ⁶	Uniformity	0.5191	F ₃ ¹³	Energy GLCM at 45°	0.5112
F ₂ ¹	Entropy Background	0.9736	F ₃ ¹⁴	Homogeneity GLCM at 45°	0.9937
F ₂ ²	Area (Shape)	0.4897	F ₃ ¹⁵	Contrast GLCM at 90°	0.1869
F ₂ ³	Perimeter (Shape)	0.5216	F ₃ ¹⁶	Correlation GLCM at 90°	0.9920
F ₂ ⁴	Solidity (Shape)	0.5981	F ₃ ¹⁷	Energy GLCM at 90°	0.5148
F ₂ ⁵	Compactness	0.4801	F ₃ ¹⁸	Homogeneity GLCM at 90°	0.9966
F ₂ ⁶	Roundness (Shape)	1.1857	F ₃ ¹⁹	Contrast GLCM at 135°	0.3468
F ₃ ¹	Texture Mean	0.9996	F ₃ ²⁰	Correlation GLCM at 135°	0.9853
F ₃ ²	Texture global mean	0.0422	F ₃ ²¹	GLCM at Energy 135°	0.5112
F ₃ ³	Texture Entropy	0.8525	F ₃ ²²	GLCM at Homogeneity 135°	0.9937
F ₃ ⁴	Texture Uniformity	0.3916	F ₃ ²³	GLCM at Homogeneity 135°	0.9937
F ₃ ⁵	Texture. Smoothness	0.0256	F ₄ ¹	Geometric Area	0.4897
F ₃ ⁶	Texture correlation	0.9950	F ₄ ²	Geometric Perimeter	0.3900
F ₃ ⁷	Homogeneity GLCM at 0°	0.9970	F ₄ ³	Geometric Compactness	0.1364

3.2.5. Classification

The semi-supervised machine learning method is used because in this proposed method manually ROIs are selected based on the number of features selected. Here, FFBPNN has been used for this classification. Classification is principally done by making expectations based on known sample information that has been learned from training data [33]. The sample data-set is separated into three sets- 70% for the training set, 15% for the testing set and 15% for the validation set. The characterization of normal and abnormal ROIs is executed using an ANN. In ANNs the three-layered feed-forward network with the

backpropagation algorithm was used [34]-[36]. Initially, bias and weights are randomly selected for FFBPNN between -0.5 to 0.5 and -1 to 1 respectively [36]. To circulate the inputs in forwarding propagation, a log sigmoid activation function is used. In (2) shows the for logistic activation function. Here, the gradient descent method used for backpropagation for iteratively examining for several set weight and bias for minimizing the mean square error among predicted network class and a known target value of the class. Afterwards, the weight is unmoving accuracy is found. Figure 5 shows the general architecture of the ANNs.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

On the study of whether the FFBPNN can acquire separated into two autonomous sets for training and testing, each with 70% information data for training, 30% information data for testing. To disregard overfitting by the FFBPNN, the validation set is utilized during the training procedure. The training set was utilized to choose the perfect set of the arrangement between associating loads of the ANNs, while the test set was utilized to survey the performance of the prepared ANNs. To train the framework Levenberg-Marquardt function is utilized, it shows extraordinary results for in training and classification [8], [36].

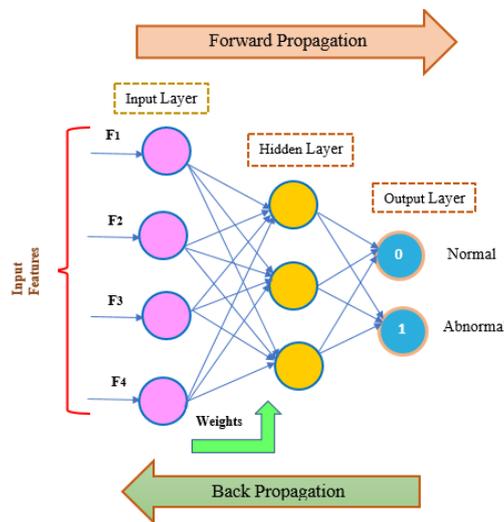


Figure 5. Architecture of ANN

4. PERFORMANCE EVALUATION MEASURES FOR FFBPNN CLASSIFIER

The experiment is simulated using MATLAB 15b. Execution of the proposed work is evaluated by considering the genuine and anticipated class. According to Table 3 confusion Matrix, TP and TN represent the number of positive and number of negative class that are classified appropriately as true positives and true negatives respectively. FP and FN are representing misclassified as false positives and false negatives respectively.

Table 3. Confusion matrix

Actual Class	Target class	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

In (3) was used for calculating the binary class classification accuracy.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{3}$$

Identification for a positive case for this binary classifier is called sensitivity/TPR/Recall. It is obtained using (4).

$$\text{Sensitivity/TPR/Recall} = \frac{TP}{TP+FN} \quad (4)$$

For negative case identification for this binary classifier is called specificity/TNR. In (5) was used for calculating this.

$$\text{Specificity/TNR} = \frac{TN}{TN+FP} \quad (5)$$

The number of positive case predictions that truly belong to the positive case define by (6).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

F- Measure gives a single score that balances both the concerns of precision and recall in one number. In (7) used for calculating F- measure.

$$F - \text{measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

In machine learning we also measure MCC of the quality of the binary class. It turns the value between -1 to +1. It is obtained by (8).

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP+FN)(TN+FP)(TP+FP)(TN+FN)}} \quad (8)$$

The specificity and false rate are identified with their relationship by the ROC curve. It can validate the effectiveness of the binary classification method, as its threshold for perception varies. The curve is formed by plotting TPR against FPR. AUC shows that the perfect outcome for the breast cancer test dataset. The AUC is near about 1 that means the accuracy of the models is high and with less accuracy will have an area nearer to 0.5. AUC is calculated as shown in (9).

$$AUC = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \quad (9)$$

5. RESULTS AND DISCUSSION

The proposed method HOFs selected 38 most important features from the mammogram images for precise binary classification. The output of this classification is in terms of normal or abnormal. In the Neural network output, normal is represented as (0 1) and abnormal as (1 0). The input to train the proposed method network is 319*38, which represent; 319 images of 38 features. The output matrix is 319*2, representing data; 319 images of 2 outputs. The output matrix format is a version of the 319 by 2 binary matrixes with 10 hidden layers. After training network, we can get the confusion matrix for this proposed method. Table 4 represents the confusion matrix of the training phase for this proposed method.

Table 4. Confusion matrix of proposed method

Actual class	Target class	
	Normal	Abnormal
Normal	207	0
Abnormal	1	111

Table 4 shows the number and percentage of exact classifications with the first two diagonal cells. The normal results show that 207 were precisely classified, corresponding to 64.9% of the total 319 mammogram images. On the other hand, the abnormal results show that 111 were precisely classified, corresponding to 34.8% of total 319 mammogram images. Only one sample is wrongly classified with an abnormal lesion, corresponding to 0.3% of the total images. The overall confusion matrix shows that, at the training stage, 99.7% of classifications are precise and 0.3% are wrong in this proposed model. So, we can say that classification accuracy achieved 99.7%. Also, we can measure the MCC parameter using the

confusion matrix is 0.9931 for this perfect binary classification. According to Table 4, MCC is calculated 0.9931 which validates of the structure being a perfect binary class for this proposed work.

ROC are beneficial tools for imaging and calculating classifiers. The performance plot is between TPR and FPR represent model classification accuracy. The ROC curve indicates that normal and abnormal samples obtained the maximum AUC as shown in Figure 6. AUC for this proposed model is 0.9975. In Figure 6(a) represent the ROC Curve for this proposed method. The F-measure score of the proposed method is 1.002. In machine learning, the main goal is to minimize the error which defines by the pattern recognition algorithm. The Mean square error of the proposed method is 2.9785×10^{-15} at epoch of 25, which means our proposed algorithm minimize the error to almost 0 as shown in Figure 6(b).

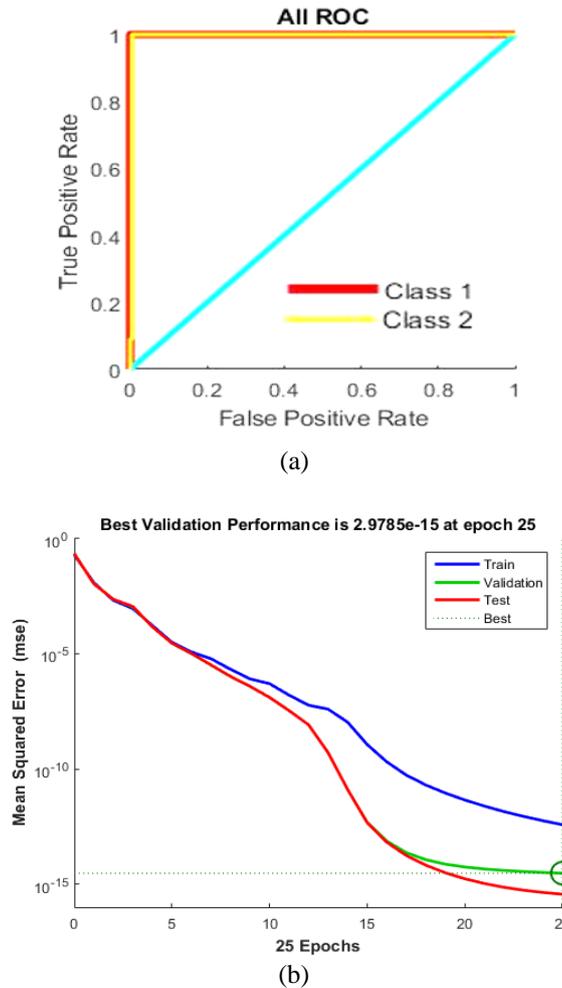
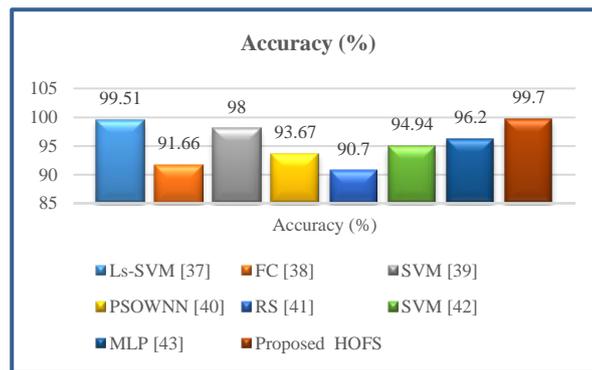
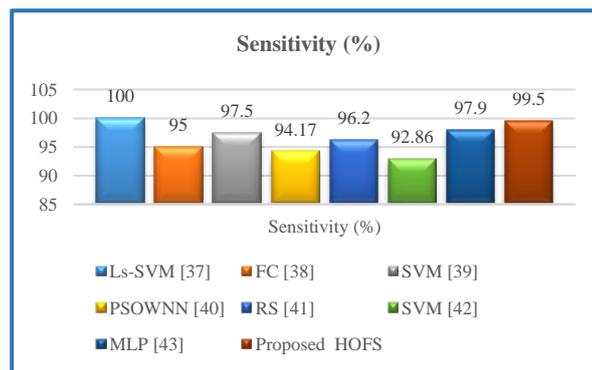


Figure 6. These figures are; (a) ROC curve for FFBPNN, (b) Performance of the training process for FFBPNN

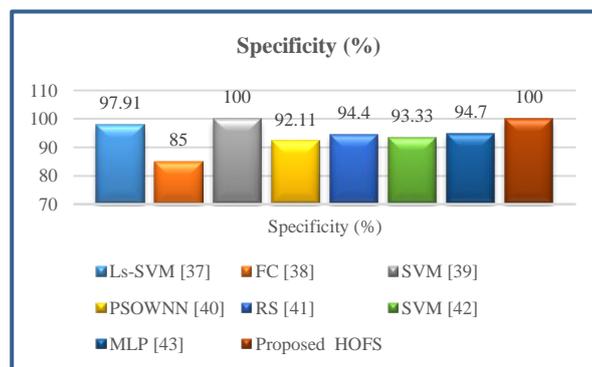
Figure 7 represents performance analysis based on HOFS and classification for breast cancer diagnosis with existing methods by performance analysis like accuracy, sensitivity and specificity. The accuracy of the HOFS method, 0.19%, 8.04%, 1.7%, 6.03%, 9%, 4.76% and 3.5% better than the existing methods Ls-SVM [37], FC [38], SVM [39], PSOWNN [40], RS [41], SVM [42], MLP [43] respectively as per figure 5(a). From figure 5(b) sensitivity of the HOFS 4.5% better than FC [38], 2% better than SVM [39], 5.33% better than PSOWNN [40], 3.3% better than RS [41], 1.6% better than MLP [43] and 6.64% better than SVM [42] models. The specificity of HOFS, 2.09%, 15%, 7.89, 5.6%, 6.67%, 5.3% better than the existing method Ls-SVM [37], FC [38], PSOWNN [40], RS [41], SVM [42], MLP [43] respectively. Therefore, from the analysis of the proposed HOFS method is higher as the above current methods, in relations of all performance measures for breast cancer analysis.



(a)



(b)



(c)

Figure 7. Performance analysis based on HOFS and classification with existing methods for classification of breast cancer, (a) Accuracy, (b) Sensitivity, (c) Specificity

Figure 8 shows that the effects of feature selection in the proposed method HOFS. Accuracy, sensitivity, specificity, F-measure, AUC and Precision are better than the without feature selection 34.2%, 2.5%, 92%, 62%, 48.2% and 34% respectively.

The results are compared with the latest techniques. In [36] they are using 251 mammogram images from MIAS database, selected 33 feature and get the accuracy 96% using FFBPNN with 50 hidden neurons, in [13] they have selected only 6 feature and achieved accuracy of 99.3% using MIAS database, in [15] they are using only 30 selected features and get the accuracy 95.7% for 70 mammogram images from MIAS database and 96.02% for 320 mammogram images from DDSM database. In our proposed HOFS model we are using 319 mammogram images from MIAS database and selected most valuable 38 features and achieves best accuracy of 99.7%, using FFBPNN with only 10 hidden neurons. So, our proposed feature selection

method is more superior as compare to some recent work in terms of accuracy, with more feature's selection and less hidden layers.

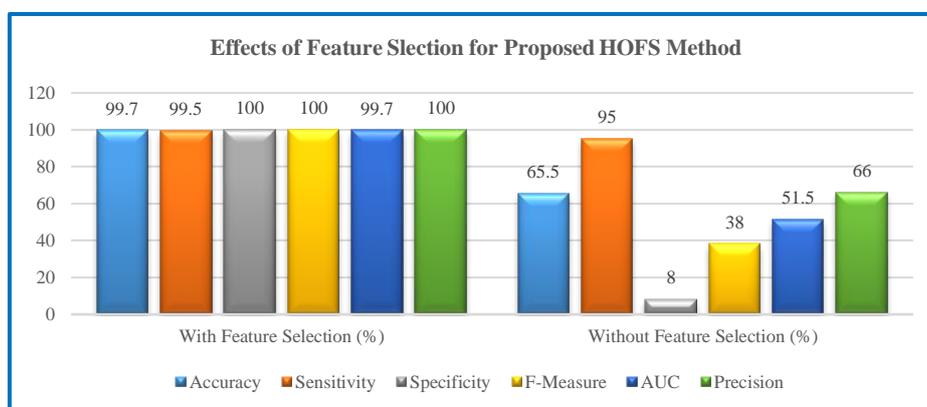


Figure 8. Effects on feature selection to improved artificial neural network for breast cancer analysis

6. CONCLUSION

In the proposed work, classification of the ROIs to identify the breast abnormal lesion. Through visual examination, our technique is effective in sectioning the abnormal part of mammogram images. In the recent existing methods classification is mostly done without feature selection process and without pre-processing of ROIs, which outcomes in the huge and redundant database. In the proposed method feature selection technique is used before classification to reduce the computation time and to increase the classification efficiency. Consideration of the complete breast region for examination is also a time-consuming process as it is a very large tissue. When features are extracted from the ROIs in CAD systems, and then its characteristics highly affect the system's efficiency. Here, from feature extraction size of the feature vector is found very large, for that a novel hybrid optimum feature selection technique is used for selecting most valuable feature to reduce the true-negative and false-negative rates and gives relatively high accuracy of 99.7%, with a 99.5% sensitivity and 100% specificity for breast cancer classification by FFBPNN. The main objective of this study was to change a fully CAD system to recognize and distinguish the normal and abnormal breast lesion by combining the mammogram images with the experimental information of breast structure. In future work, the point is to accomplish exact malignancy recognition with huge datasets.

REFERENCES

- [1] Statistics of Breast Cancer in India. Global comparison, 2018, Retrieved from <http://www.breastcancerindia.net>.
- [2] E. Fernandez, J.R. Gonzalez, J.M. Borrás, V. Moreno, V. Sanchez, M. Peris, "Recent decline in cancer mortality in Catalonia (Spain) A joint point regression analysis," *European Journal of Cancer* 37 pp. 2222-2228, 2001, doi: 10.1016/s0959-8049(01)00279-9.
- [3] Rodriguez C, Plasencia A, Schroeder DG, "Predictive factors of enrollment and adherence in a breast cancer screening program in Barcelona (Spain)," *Soc Sci Med.*, vol. 40, no.8 pp. 1155-60, 1995, doi: 10.1016/0277-9536(94)00184-u.
- [4] Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin.* vol.63, no. 1, pp.11-30, 2013, doi: 10.3322/caac.21166.
- [5] Yasmin M, Sharif M, Mohsin S, "Survey Paper on Diagnosis of Breast Cancer Using Image Processing Techniques," *Res J Recent Sciences*, vol. 2, no.10, pp. 88-98, 2013, www.isca.in, www.isca.me.
- [6] A. M. Abdel-zaher and A. M. Eldeib, "Breast Cancer Classification using Deep Belief Networks," *Expert Systems with Applications* vol. 46, pp 139-144, 2016, doi: 10.1016/j.eswa.2015.10.015.
- [7] Susama Bagchi, Kim Gaik Tay, Audrey Huong, Sanjoy Kumar Debnath, "Image processing and machine learning techniques used in computer-aided detection system for mammogram screening-A review," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3, pp. 2336-2348 2020, doi: 10.11591/ijece. v10i3.pp2336-2348.
- [8] Khosrowshahi, F., "Innovation in artificial neural network learning: Learn-On-Demand methodology," *Automation in Construction*, vol. 20, no.8: pp. 1204-1210, 2011, doi: 10.1016/j.autcon.2011.05.004

- [9] Cheok, C.Y., *et al.*, "Optimization of total phenolic content extracted from *Garcinia mangostana* Linn. Hull using response surface methodology versus artificial neural network," *Industrial Crops and Products*, vol. 40: pp. 247-253, 2012, doi: 10.1016/j.indcrop.2012.03.019.
- [10] S.-C. B. Lo, H.-P. Chan, J.-S. Lin, H. Li, M. T. Freedman, and S. K. Mun, "Artificial convolution neural network for medical image pattern recognition," *Neural Networks*, vol. 8, no. 7-8, pp. 1201-1214, 1995, doi: 10.1016/0893-6080(95)00061-5.
- [11] H. Tang, K. C. Tan, and Z. Yi, "Competitive neural networks for image segmentation," *Studies in computational intelligence*, vol. 53, pp. 129-144, 2007, doi: 10.1016/j.jksuci.2018.01.004.
- [12] Shankar Thawkar, RanjanaIngolikar, "Classification of masses in digital mammograms using Biogeography-based optimization technique," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 10, pp.1140-1148, 2020, doi: 10.1016/j.jksuci.2018.01.004.
- [13] Nadeem Tariq, "Breast Cancer Detection using Artificial Neural Networks," *Journal of Molecular Biomarkers J & Diagnosis*, vol. 9, no. 1, 2017, doi: 10.4172/2155-9929.1000371.
- [14] Amit Kamra, Poonam Sood and Akshay Girdhar, "Investigation on ROI Size and Location to Classify," *Int. J. Biomedical Engineering and Technology*, vol. 29, no. 1, pp. 68-83, 2019, doi: 10.1504/IJBET.2019.096881.
- [15] Weiying Xie, Yunsong Li, Yide Ma, "Breast Mass Classification in Digital Mammography Based on Extreme Machine Learning," *Neurocomputing*, vol.173, no. 3, pp. 930-941, 2016, doi: 10.1016/j.neucom.2015.08.048.
- [16] Anuj Kumar Singh and Bhupendra Gupta, "A Novel Approach for Breast Cancer Detection and Segmentation in a Mammogram," *Procedia Computer Science*, vol.54, pp. 676-682, 2015, doi: 10.1016/j.procs.2015.06.079
- [17] S. Punitha, A. Amuthan, K.Suresh Joseph, "Benign and Malignant Breast Cancer Segmentation Using Optimized Region Growing Technique," *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 348-358, 2018, doi: 10.1016/j.fcij.2018.10.005.
- [18] Rouhi R, Jafari M, Kasaei S, Keshavarzian P, "Benign and Malignant Breast Tumors Classification Based on Region Growing and CNN Segmentation," *Expert System with Applications*, vol. 42, no. 3, pp. 990-1002, 2015, doi:10.1016/j.eswa.2014.09.020.
- [19] Arnau Oliver, Joan Marti, Robert Marti, AnnaBosch, and Jordi Freixenet, "A New Approach to the Classification of Mammographic Masses and Normal Breast Tissue," 2006, *18th International Conference on Pattern Recognition. IEEE*, vol. 4, pp. 707-710, doi:10.1109/ICPR.2006.113.
- [20] K. Bovis and S. Singh, "Detection of masses in mammograms using texture features," *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Barcelona, Spain, 2000, pp. 267-270 vol. 2, doi: 10.1109/ICPR.2000.906064.
- [21] Khuzi AM, Besar R, Zaki, "Identification of Masses in Digital Mammogram Using Gray Level Co-Occurrence Matrices," *Biomed. Imaging Intervention Journal*, vol. 5, no. 3, 2009, doi: http://dx.doi.org/10.2349/bij.5.3.e17.
- [22] M.M. Mehdy, P. Y. Ng, E. F. Shair, N.L. Md Saleh, C. Gomes, "Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer," *Computational and Mathematical Methods in Medicine*, vol. 2017, 2017, doi:10.1155/2017/2610628.
- [23] Mini MIAS data base peipa.essex.ac.uk/info/mias.htm.
- [24] Mariam Biltawi, Nijad Al-Najdawi, Sara Tedmori, "Mammogram Enhancement and Segmentation Methods: Classification, Analysis, and Evaluation," 2012, *13th International Conference on Advances in Computing and Information Technology* Dec.10-13, <https://www.researchgate.net/publication/248392529>.
- [25] Shivhare E, Saxena V, "Breast cancer diagnosis from mammographic images using optimized feature selection and neural network architecture," *Int J Imaging Syst Technol*, pp. 1-17, 2020, doi: 10.1002/ima. 22467.
- [26] Md Akizur Rahman and Ravie Chandren Muniyandi, "An Enhancement in Cancer Classification Accuracy Using a Two-Step Feature Selection Method Based on Artificial Neural Networks with 15 Neurons," *Symmetry*, vol. 12, no.2, 271, 2020, doi:10.3390/sym12020271.
- [27] Y. Ireaneus Anna Rejani, S.Thamarai Selvi, "Breast Cancer Detection Using Multilevel Thresholding," *International Journal of Computer Science and Information Security*, vol.6, no.1, pp. 111-115, 2009, <https://arxiv.org/abs/0911.0490v1>.
- [28] Mohammadhasan Salimian, Abdalhossein Rezai, Seyedeh Shahrbanoo Falahieh Hamidpour, Farzad Khajeh-Khalili, "Effective Features in Thermal Images for Breast Cancer Detection," 2019, *2nd National Conference on New Technologies in Electrical and Computer Engineering*, pp. 1-7.
- [29] R.M. Haralick, K. Shanmugam, and I.H. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp.610-621, 1973, doi: 10.1109/TSMC.1973.4309314.
- [30] Wang, J., Nishikawa, R. M., and Yang, Y., "Global Detection Approach for Clustered Micro calcifications In Mammograms Using a Deep Learning Network," *Journal of Medical Imaging*, vol. 4, no. 2, 2017, doi: 10.1117/1.JMI.4.2.024501.
- [31] Shofwatul Uyun, Lina Choridah, "Feature Selection Mammogram based on Breast Cancer Mining," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 1, pp. 60-69, 2018, doi: 10.11591/ijece. v8i1.pp60-69.
- [32] Rahman, M.A.; Muniyandi, R.C, "Feature selection from colon cancer dataset for cancer classification using artificial neural network," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4-2, pp.1387-1393,2018, doi: 10.18517/ijaseit.8.4-2.6790.
- [33] Patel, J.J., & Hadia, S.K, "A Review Paper on Digital Mammography for Detection of Breast Cancer," *International Journal of Research in Advent Technology*, vol. 6, no. 5, 2018, www.ijrat.org.
- [34] Rumelhart DE, McClelland JL, the PDP Research Group, "Parallel distributed processing," *The MIT Press, Cambridge, Massachusetts* vol. 1, 1986.

- [35] Wang, J., Nishikawa, R. M., and Yang, Y, "Global Detection Approach for Clustered Micro calcifications In Mammograms Using a Deep Learning Network," *Journal of Medical Imaging*, vol. 4. no. 2, 2017, doi.org/10.1117/1. JMI.4.2.024501.
- [36] K.U. Sheba, S. Gladston Raj, "An approach for automatic lesion detection in mammograms," *Cogent Engineering*, V 5: 1444320, 2018, doi: 10.1080/23311916.2018.1444320.
- [37] MF Akay, "Support Vector Machines Combined With Feature Selection For Breast Cancer Diagnosis" *Expert Systems with Applications*, vol. 36, no. 2, Part 2, pp. 3240-3247, 2008, doi.org/10.1016/j.eswa.2008.01.009.
- [38] M.A. Saleem Durai, Kanna A, Narayan SN, "Fuzzy Classification Model Assisted by Intensity Based Approach and Diagnosis," *International Journal of Advanced Research in Computer Science*, vol. 1, no. 3, pp. 398-402, 2010, www.ijarcs.info.
- [39] Monica Jenefer and Cyrilraj, "An Efficient Image Processing Methods for Mammogram Breast Cancer Detection," *Journal of theoretical and applied information technology*, vol. 69, no. 1, pp. 32-39, 2014, www.jatit.org.
- [40] 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*, vol.49, pp. 45-52, 2014, doi:10.1016/j.jbi.2014.01.010.
- [41] Nijad Al-Najwadi, Mariam Biltawi and Sara Tedmori, "Mammogram Image Visual Enhancement, Mass Segmentation and Classification," *Applied Soft Computing*, vol. 35, pp. 175-185, 2015, doi: 10.1016/j.asoc.2015.06.029.
- [42] De Oliveira Martins L., Junior G.B., da Silva E.C., Silva A.C., de Paiva A.C., (2007) Classification of Breast Tissues in Mammogram Images Using Ripley's K Function and Support Vector Machine. In: Kamel M., Campilho A. (Eds) *Image Analysis and Recognition. ICIAR 2007*. Lecture Notes in Computer Science, vol 4633. Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-74260-9_80.
- [43] Hanung Adi Nugroho, Hesti Khuzaimah Nurul Yusufiyah, Teguh Bharata Adji, Widhia K. Z Oktoeberza, "Shape analysis for classification of breast nodules on digital ultrasound images," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 2, pp. 837-844, 2019, doi: 10.11591/ijeecs. v13.i2.

BIOGRAPHIES OF AUTHORS



Ms. Jalpa J. Patel pursued Bachelor of Engineering from Bhavnagar University, Bhavnagar in 2007 and Master of Engineering from Sardar Patel University in year 2009. She is currently pursuing Ph.D. from V.T. Patel Department of Electronics & Communication Engineering, Charotar University of Science & Technology, since 2017. She is a life time member of IETE since 2010. She has published more than 8 research papers in reputed international journals and conferences and it's also available online. Her main research work focuses on Bio medical image processing. She has 8 years of teaching experience 1.6 years of Industry experience and 1.6 years of reseach experience.



Dr. S K Hadia is an associate professor at Gujarat Technological University (GTU), Ahemdabad, and Gujarat, India. He has completed his Ph.D. from Charotar University of Science and Technology in year 2016. His main research work focuses on Optical communication computer network, Image processing. He has more than 15 years of teaching experience. He has published more than 11 research papers in reputed high impact international journals and conferences.