Classification of mammogram using two-dimensional discrete orthonormal S-transform for breast cancer detection

Shradhananda Beura ✉, *Banshidhar Majhi, Ratnakar Dash, Susnata Roy*

Department of Computer Science and Engineering, Pattern Recognition Laboratory, National Institute of Technology, Rourkela 769008, India
✉ *E-mail: beura.shradhananda@gmail.com*

Published in Healthcare Technology Letters; Received on 22nd May 2014; Accepted on 28th January 2015

An efficient approach for classification of mammograms for detection of breast cancer is presented. The approach utilises the two-dimensional discrete orthonormal S-transform (DOST) to extract the coefficients from the digital mammograms. A feature selection algorithm based the on null-hypothesis test with statistical 'two-sample *t*-test' method has been suggested to select most significant coefficients from a large number of DOST coefficients. The selected coefficients are used as features in the classification of mammographic images as benign or malignant. This scheme utilises an AdaBoost algorithm with random forest as its base classifier. Two standard databases Mammographic Image Analysis Society (MIAS) and Digital Database for Screening Mammography (DDSM) are used for the validation of the proposed scheme. Simulation results show an optimal classification performance with respect to accuracies of 98.3 and 98.8% and AUC (receiver operating characteristic) values of 0.9985 and 0.9992 for MIAS and DDSM, respectively. Comparative analysis shows that the proposed scheme outperforms its competent schemes.

1. Introduction: Breast cancer is currently one of the major reasons of an increased death rate among women. Early detection through periodic screening improves the chance of recovery from breast cancer. For a reliable early detection, mammography is an effective method in which digital mammograms are analysed [1]. Digital mammograms are the scanned X-ray images of breasts. Interpretation of mammograms is a very important task for radiologists as they refer patients for biopsy. However, interpretation of mammograms varies among radiologists as it depends on training and experience. This leads to different judgments by different radiologists. It has been observed that 60–90% of initially suspected malignant lesions by radiologists were found to be benign later [2]. Therefore, avoidance of misinterpretation is highly desirable. Currently, computer-aided diagnosis is a very popular and efficient method that analyses the digital mammograms and helps radiologists in mammogram interpretation to detect the suspicious lesions as well as their type. Regarding this responsibility, one important step is to extract a set of significant features from the mammographic images that can distinguish the benign lesions from malignant ones. Different techniques and methods have been studied for the extraction of features and the classification of mammograms into benign and malignant classes.

Liu *et al.* [3] achieved an accuracy of 84.2% to classify the benign and malignant mammograms by using a set of wavelet based statistical features and a binary tree as the classifier. Pereira *et al.* [4] proposed a method in which a spatial grey level dependence matrix of wavelet transformed mammograms was used to derive the texture features. These texture features were utilised to classify the mammograms as benign or malignant with the help of a non-parametric K-nearest neighbours (K-NNs) classifier. In their method, the performance index values of AUC = 0.617 and 0.607 have been achieved for discriminating the masses and micro-calcifications, respectively. Verma *et al.* [5] used BI-RADS descriptor features to classify the mammograms with their proposed soft clustered based direct learning classifier and achieved an accuracy of 97.5%. Fraschini [6] used discrete wavelet transform and neural network to classify the mammograms. The performance index value was AUC of 0.91 in the ROC curve. Buciu and Gacsadi [7] achieved AUC value of 0.78 for the classification of benign-malignant mammograms. They used Gabor wavelets with principal component analysis for reduction in the dimension of directional features with the help of a support vector machine (SVM) as a classifier. Prathibha and Sadasivam [8] obtained a classification accuracy of 90.65% for benign and malignant mammogram characterisation by using a combination of wavelet and SVM. Xiaoming *et al.* developed a method using an SVM-based recursive feature elimination (SVM-RFE) procedure with a normalised mutual information feature selection (NMIFS) and achieved a performance measure AUC of 0.9615 in the mammogram classification [9]. They used geometry features, relative gradient orientation-based features, and texture feature like grey-level co-occurrence matrix in their proposed scheme. Zanchetta *et al.* [10] used discrete wavelet transform for the extraction of features and a polynomial classifier to classify the benign-malignant mammograms. They achieved the performance measure AUC of 0.95. Görgel *et al.* [11] achieved 91.67% classification accuracy for benign and malignant mammogram classification using spherical wavelet transform (SWT) for extraction of features and SVM as the classifier. In their proposed method, a local seed region growing algorithm was used to detect the region-of-interests (ROIs) of mammograms. Kumar and Balakrishnan [12] developed a method based on the combination of DWT and stochastic neighbour embedding technique for benign and malignant mammogram classification. They used stochastic neighbour embedding technique to reduce wavelet coefficients of mammograms and achieved classification accuracy of 90.10% with the help of an SVM classifier. Ganesan *et al.* [13] proposed a one-class classification method to classify the mammographic images as benign or malignant. A trace transform functional was used for extraction of features from the mammograms and 92.48% of accuracy rate was obtained by using a Gaussian mixture model (GMM).

In this Letter, we propose a scheme in which a two-dimensional (2D) discrete orthonormal S-transform (DOST) has been used to extract features from mammographic images. A feature selection algorithm has been suggested using null-hypothesis test with 'two-sample *t*-test' method to select significant features from the available set of extracted features. A combination of the AdaBoost algorithm with random forest (AdaBoost-RF) classifier is used to classify the mammogram as benign or malignant using the selected

doi: 10.1049/htl.2014.0108

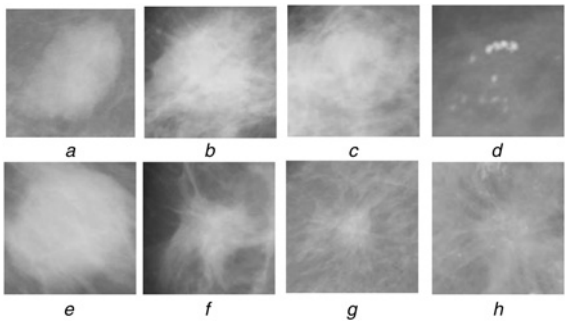

Figure 1 *Extracted ROIs from different mammographic images (source: MIAS database)*
a–d Benign type
e–h Malignant type

features. The obtained results are compared with the existing methods to validate the efficacy of the proposed scheme.

2. Proposed method: The proposed method consists of three important phases: feature extraction, selection and classification. Conventional cropping operations are employed to select the region-of-interests (ROIs) as shown in Fig. 1. These ROIs become inputs to our suggested scheme. The overall block diagram of the proposed scheme is shown in Fig. 2.

2.1. Feature extraction: The scheme uses a 2D DOST which is a multi-scale technique to extract the pixel-by-pixel texture features of a mammographic image. DOST is based on the S-transform, which is a time–frequency representation closely related to continuous wavelet transform [14]. The S-transform is advantageous for the analysis of mammographic images as it preserves the phase information using linear frequency scaling. However, the major limitation of S-transform is its high time and space complexity because of its redundant nature. To remove these limitations, DOST uses an orthonormal set of basis functions. Therefore DOST has less computational and storage complexity when compared to S-transform. DOST of a mammographic image, $f(x, y)$ of size $N \times N$ can be obtained using a dyadic sampling scheme given by the following steps:

1. Perform 2D Fourier transform (2D-FT) on the image $f(x, y)$ of size $N \times N$ to obtain Fourier samples, $F(u, v) \leftarrow$ 2D-FT$[f(x, y)]$.
2. Partition $F(u, v)$ and determine the number of points in that partition.

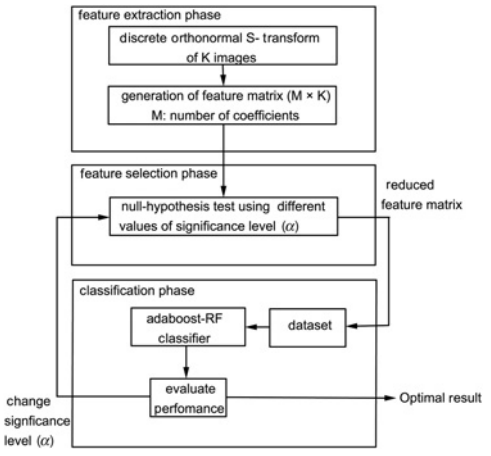

Figure 2 *Block diagram of proposed scheme for classification of mammograms*

3. Compute the square root of the number of points and multiply it with $F(u, v)$ to obtain a result.
4. Apply an inverse 2D-FT on the result to obtain the DOST description of the image $f(x, y)$, which is termed as voice image and given by

$$S(x', y', \upsilon_x, \upsilon_y) = \frac{1}{\sqrt{2^{p_x+p_y-2}}} \times \sum_{u=-2^{p_x-2}}^{2^{p_x-2}-1} \sum_{v=-2^{p_y-2}}^{2^{p_y-2}-1} F(u+\upsilon_x, v+\upsilon_y) \times e^{2\pi i\left((ux'/2^{p_x-1})+(vy'/2^{p_y-1})\right)} \quad (1)$$

where $\upsilon_x = 2^{p_x-1} + 2^{p_x-2}$ and $\upsilon_y = 2^{p_y-1} + 2^{p_y-2}$ are horizontal and vertical voice frequencies.
5. A rectangular voice image is obtained having $2^{p_x-1} \times 2^{p_y-1}$ points same as in the original image.

In DOST, each pixel $p(x, y)$ within the image gives voice frequencies (υ_x, υ_y) with $2^{p_x-1} \times 2^{p_y-1}$ bandwidth. Subsequently the pixel-wise local spatial frequency description in DOST is computed as

1. Select an arbitrary pixel at coordinate (x, y) within the image.
2. Compute the value of the voice image S, $\forall(\upsilon_x, \upsilon_y)$ in the frequency order (p_x, p_y) of the location (x, y) at $S[x/N \times 2^{p_x-1}, y/N \times 2^{p_y-1}]$.
3. Build a local spatial frequency domain having size $2\log_2 N \times 2\log_2 N$ by iterating over all values (p_x, p_y) for each pixel of the image.

The frequency domain contains positive and negative components from DC, $(\upsilon_x, \upsilon_y) = (0, 0)$ to the Nyquist frequency N_f, $(\upsilon_x, \upsilon_y) = (N/2, N/2)$. Thus, all the components in the frequency domain are mapped to the M-space frequency coefficients. In this way, a $N \times N$ ROI generates $N \times N$ DOST coefficients and each coefficient is included in the feature vector (FV). Combination of K FVs is represented in a feature matrix FM. The details feature extraction process is described in Algorithm 1. The FM becomes an input to the feature selection phase.

Algorithm 1 Feature Extraction

Require: K : Total number of ROIs taken for the experiment
Ensure: $FM[1{:}M, 1{:}K]$: Feature matrix
R: Total number of reduced features
Function *dost*() computes DOST coefficients of ROIs. Function *resize*() sets the dimension of each ROI as per required.
1: Create an empty matrix $CM[1{:}N, 1{:}N]$ and an empty vector FV {CM is used as DOST coefficient matrix and FV is used as feature vector
2: Initialise N in terms of pixel, $i \leftarrow 1$
3: $M \leftarrow N \times N$ {Total number of features is to be extracted from a ROI}
4: **for** $k \leftarrow 1$ to K **do**
5: Get ROI_k
6: $ROI_k \leftarrow resize(ROI_k, N)$
7: $CM_k[1{:}N, 1{:}N] \leftarrow dost(ROI_k)$
8: **for** $p \leftarrow 1toN$ **do**
9: **for** $q \leftarrow 1toN$ **do**
10: $FV_k[i, 1] \leftarrow CM_k[p, q]$
11: $i \leftarrow i + 1$
12: Reset $i \leftarrow 1$
13: **for** $m \leftarrow 1toM$ **do**
14: $FM[m, k] \leftarrow FV_k[m, 1]$

2.2. Feature selection and classification: In the feature selection phase, an optimal set of relevant features is selected from the extracted feature matrix. Here, a statistical null-hypothesis test using 'two-sample *t*-test' method [15] is used to select features. The null-hypothesis test is carried out on two normally distributed populations of samples, say BE and MA, containing benign and malignant feature data, respectively. The test decision specifies whether the null-hypothesis is correct or incorrect, which in turn triggers the data from two populations and determines whether they are significantly different or not. The incorrect null-hypothesis is rejected and specifies that the data from two populations are significantly different from each other and independent. Whereas, a correct null-hypothesis is not rejected and there is no significant difference between the data from two populations.

Let instances of two populations $b_i \in \text{BE}$, $i = 1, 2, \ldots, n_1$ and $m_j \in \text{MA}$, $j = 1, 2, \ldots, n_2$ be FVs. The corresponding means and standard deviation of two populations BE and MA are μ_{BE}, μ_{MA}, and σ_{BE} and σ_{MA}, respectively. Now, the null-hypothesis test is performed in the following steps:

1. Specify the desired value of significance level (α) between 0 and 1. The significance level is the probability of null-hypothesis being incorrect.
2. Compute the *t*-test

$$t = \left(\left|\mu_{\text{BE}} - \mu_{\text{MA}}\right|\right) \Big/ \sqrt{\left(\frac{\sigma_{\text{BE}}^2}{n_1} + \frac{\sigma_{\text{MA}}^2}{n_2}\right)} \tag{2}$$

3. Calculate the degrees of freedom

$$d = \left(\frac{\sigma_{\text{BE}}^2}{n_1} + \frac{\sigma_{\text{MA}}^2}{n_2}\right)^2 \Big/ \left(\frac{\sigma_{\text{BE}}^4}{n_1^2(n_1 - 1)} + \frac{\sigma_{\text{MA}}^4}{n_2^2(n_2 - 1)}\right) \tag{3}$$

4. Compute the *p*-value using the cumulative distributed function of *t*-test statistics

$$p = \int_{-\infty}^{t} \frac{\Gamma\left((d+1)/2\right)}{\sqrt{\pi \times d} \times \Gamma(d/2)} \times \left(1 + \frac{t^2}{d}\right)^{-((d+1)/2)} \tag{4}$$

where Γ is a Gamma function. The *p*-value is the probability of the *t*-test with degrees of freedom d given that the null-hypothesis is correct.
5. Set the decision value for the null-hypothesis test

$$h = \begin{cases} 1, & \text{if } p \le \alpha \\ 0, & \text{if } p > \alpha \end{cases} \tag{5}$$

6. For $h = 1$, the null-hypothesis is incorrect and rejected for the specified value of α.

We have taken the label values 0 and 1 for representing the sample as benign and malignant classes, respectively. The *target* vector contains the label values of all ROI samples that are used in the scheme. With the help of the *target* vector, the two populations BE and MA are generated. From the null-hypothesis testing, the returned decision value is a vector, defined as, $h_m \in \{0, 1\}$, $m = 1, 2, \ldots, M$, where M represents the total number of extracted features. Then, a feature $f_m \in \text{FM}$ is to be selected as a relevant one, if and only if h_m equals 1. Thus, all the selected relevant features are collected from the feature matrix (FM) to form a reduced FM (RFM) for K number of ROIs. The total number of reduced feature(s) denoted as R is decided according to the value of α specified in the hypothesis testing. Further more, a training dataset X is created for K number of ROIs using the RFM and *target* vector, which is used in the classifier to design an effective classifier model. Algorithm 2 describes the feature selection process in details.

Algorithm 2 Feature Extraction

Require: $FM[1{:}M, 1{:}K]$, $target[1{:}K]$
α : Significance level
Ensure: $RFM[1{:}R, 1{:}K]$: Reduced feature matrix
R: Total number of reduced features
Function **nhtest()**s computes statistical null-hypothesis decision value using two vectors at different values of α using *two-sample t-test* method.
Create two empty vectors BE and MA
Initialize α with $0 < \alpha < 1$ and $i \leftarrow 1, j \leftarrow 1, l \leftarrow 1$
for $m \leftarrow 1$ to M
  Clear contents of vector BE and vector MA
  for $k \leftarrow 1$ to K **do**
  if $target[i] = 1$ **then**
    $target[k] = 1$
    $MA[1, i] \leftarrow FM[m, k]$
    $i \leftarrow i + 1$
  else
    $BE[1, j] \leftarrow FM[m, k]$
    $j \leftarrow j + 1$
Reset $i \leftarrow 1$ and $i \leftarrow 1$
$h[i] =$ **nhtest**(BE, MA, α)
  if $h[m] = 1$ **then**
  for $k \leftarrow 1$ to K **do**
    $RFM[l, k] \leftarrow FM[m, k]$
  $l \leftarrow l + 1$

An AdaBoost algorithm has been used with the random forest classifier as a base or weak learner for the benign-malignant mammogram classification. Random forest is an ensemble classification technique developed by Breiman, which shows an adequate performance with respect to the accuracy rate [16]. Random forest uses the bagging technique in which each individual classifier is built by bootstrap input samples. In this method, a number of features randomly selected and used to make a decision at the node split. The error rate of the random forest depends on the number of features used in the decision node. This method is the combination of many tree predictors in which each tree depends on the values of random vector sampled independently and with the same distribution for all the trees in the forest. Using the training set and random vector, a tree is built. After a number of trees are generated, a classifier hypothesis is generated based on the voting for the class.

AdaBoost algorithm is the most popular version of the boosting procedure. It has the highest flexibility for adding many weak classifiers having high error rates to generate a combined hypothesis whose training error rate is small [17–19]. For the two class classification problem, the AdaBoost algorithm takes the training dataset $X = (\boldsymbol{x}_i, y_i)$, $i = 1, 2, \ldots, K$, where $\boldsymbol{x}_i$ is a vector containing features and $y_i \in \{0, 1\}$ is the label of class. As mentioned earlier, the label value 0 and 1 represent negative class (benign) and positive class (malignant) ROIs, respectively. In the algorithm, each training example is assigned by a weight that determines the probability of that example being selected for the training set for a classifier. Initially, the algorithm gives equal weights to all training examples. Then, the weak classifier is trained repeatedly on the weighted versions of training examples for a series of rounds. In each round, a weak hypothesis of lower error rate is generated, and at the same time, the weights of misclassified examples are increased. The examples with higher weights are selected to train the classifier in the next round. Thus, the final decision is obtained by the linear combination of the weak hypothesis generated in each round. The detailed

doi: 10.1049/htl.2014.0108

description of the AdaBoost procedure with random forest classifier is given in Algorithm 3.

Algorithm 3 Classification with AdaBoost-RF

Require: Dataset having K instances, $X = x_i$ for $i = 1, 2, ..., K$ with levels $y_i \in \{0, 1\}$
Random forest classifier as weak learner
N: Total number of iteration
T: Total number of trees
Ensure: *classifier_decision*
1: Initialise weight $W_1(i) \leftarrow 1/K, \ \forall i$
2: **for** $n \leftarrow 1$ to N **do**
3: **for** $t \leftarrow 1$ to T **do**
4: Generate a vector V_t with $W_n(i)$
5: $X_t \leftarrow bootstrap(X)$
6: $ctree_t \leftarrow buildtree(X_t, V_t)$
7: **return** hypothesis h
8: Obtain class hypothesis, $h_n(x_i) \rightarrow y_i \in \{0, 1\}$
9: Compute the error of $h_n(x_i)$,

$$error_n \leftarrow \sum_{i:h_n(x_i)\neq y_i} W_n(i)$$

10: Set a constant $c_n \leftarrow \frac{1}{2}\ln\left(\frac{1 - error_n}{error_n}\right)$
11: $W_{n+1}(i) \leftarrow \frac{W_n(i)}{Z_n} e^{-c_n y_i h_n(x_i)}$ {Z_n is a normalising constant}
12: $h_{final}(x) \leftarrow \sum_{n=1}^{N} c_n h_n(x)$
13: $classifier\_decision \leftarrow sign[h_{final}(x)]$

The performance of the classification algorithm is evaluated with the help of confusion matrix, ROC curve with AUC score, and other parameters like F-measure and Matthews correlation coefficient (MCC). A confusion matrix is a table that provides information about the predicted and actual class classification performed by the classifier. The confusion matrix for two classes (benign and malignant) and performance measures are given in Tables 1 and 2, respectively.

The TPR (true positive rate) and FPR (false positive rate) are two important measures for performance evaluation. The TPR calculates malignant ROIs correctly classified out of the total number of malignant ROIs. The FPR parameter calculates benign ROIs incorrectly classified out of the total number of benign ROIs. The F-measure and MCC play an important role in the quality evaluation of binary classification. The F-measure is computed as the harmonic mean of 'precision' and 'recall' and given by

$$F-\text{measure} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (6)$$

The MCC is a correlation coefficient between the observed and predicted classification and given by

$$\text{MCC} = \frac{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})}{(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TP} + \text{FP})(\text{TN} + \text{FN})} \quad (7)$$

The F-measure ranges from 0 to + 1 and MCC ranges from −1 to +1. Larger values of both F-measure and MCC indicate higher classification quality. The evaluation of a classifier performance can also be achieved by means of receiver operating characteristics (ROC) curve. ROC curve is represented in a 2D graph which plots TPR against FPR. The ROC curve has an important index value known as the area under the curve (AUC), which determines the classifier's performance. The AUC of value 1.0 is an ideal performance of the classifier.

Table 1 Confusion matrix for two classes

Actual class	Predicted class	
	Positive	Negative
positive	TP (true positive)	FN (false negative)
negative	FP (false positive)	TN (true negative)

Table 2 Different measures of classification performance

Measure	Definition
TPR or recall	TP/(TP + FN)
FPR	FP/(FP + TN)
precision	TP/(TP + FP)
ACC	(TP + TN/K)

TPR: true positive rate, FPR: false positive rate, K: total number of samples, ACC: accuracy.

3. Experimental results and discussion: In order to validate the proposed scheme, we have taken 115 digital mammographic images from Mammographic Image Analysis Society (MIAS) database [20] and 250 images from Digital Database for Screening Mammography (DDSM), which are collected from Image Retrieval in Medical Applications (IRMA) project [21]. All these images belong to two classes of abnormality such as benign and malignant. Both MIAS and DDSM databases provide appropriate information based on the background tissues and the class of abnormalities. Out of 115 MIAS images, 64 images are benign type and 51 images are of malignant type. Similarly, the number of benign and malignant DDSM mammograms are 121 and 129, respectively, collected from the IRMA project. In this work, all the mammographic ROI have been cropped to 128 × 128 pixels in size, and are used in the feature extraction experiment. In the feature extraction phase (Algorithm 1), 128 × 128 DOST coefficients are extracted from each ROI and a feature matrix is built by keeping all corresponding coefficients of each ROI in rows and ROI indices in columns.

Next, the significant features are selected using the feature selection algorithm (Algorithm 2). The RFMs are generated by using different values of significance level (α). Using these RFMs and the class vector (*target*), a number of datasets are generated and used in the classifier. We have employed a ten-fold cross-validation technique for each experiment for a number of rounds. In the ten-fold cross-validation experiment, the whole dataset is partitioned into ten number of folds. In each round, nine folds are combined to form one set and the remaining fold forms another set. Thus, two disjoint sets are formed containing 10 and 90% data that are used separately for training and validation process, respectively. This process is repeated ten times with random selection of training and testing data by the classifier. For the classification job, we have taken random forests with 10, 20, 40, 80 and 100 trees, with maximum depth of two, which is used as the base learner in the AdaBoost algorithm. It has been observed that the best performance is achieved using a random forest with 20 trees. Thus, with optimal structure of the classifier, a number of datasets having various sizes are used for the classification of benign–malignant mammograms. The results are articulated in Table 3. The optimum performance has been achieved with 98.3 and 98.8% accuracies using MIAS and DDSM databases, respectively, with p-value less than $\alpha = 7 \times 10^{-4}$. The other parameters such as F-measure, MCC and AUC are also maximum at that optimal $\alpha = 7 \times 10^{-4}$. At this value of significance level, the root relative square errors (0.239 – MIAS, 0.1194 – DDSM) are also minimum than that of at other values of α as shown in Table 3. The fold-wise results in terms of confusion matrix for optimal dataset (at $\alpha = 7 \times 10^{-4}$) in ten-fold cross-validation experiment is also presented in Table 4.

Table 3 Comparative analysis at different values of α

Significance level (α)	Performance measures											
	MIAS						DDSM					
	TPR	FPR	*F*-measure	MCC	AUC	ACC, %	TPR	FPR	*F*-measure	MCC	AUC	ACC, %
6×10^{-3}	0.882	0.109	0.874	0.772	0.930	88.7	0.930	0.124	0.909	0.808	0.905	90.4
5×10^{-3}	0.922	0.031	0.940	0.894	0.979	94.8	0.953	0.124	0.921	0.834	0.916	91.6
4×10^{-3}	0.953	0.118	0.931	0.842	0.956	92.2	0.969	0.116	0.933	0.858	0.933	92.8
3×10^{-3}	0.922	0.031	0.940	0.894	0.979	94.8	0.984	0.099	0.948	0.890	0.982	94.4
2×10^{-3}	0.961	0.047	0.951	0.912	0.992	95.6	0.946	0.066	0.942	0.880	0.941	94.0
1×10^{-3}	0.941	0.016	0.960	0.930	0.994	96.5	0.977	0.041	0.969	0.936	0.996	96.8
8×10^{-4}	0.961	0.016	0.970	0.947	0.996	97.4	0.977	0.025	0.977	0.952	0.998	97.6
7×10^{-4}	0.961	0	0.980	0.965	0.998	**98.3**	0.984	0.008	0.988	0.976	0.999	**98.8**
6×10^{-4}	0.961	0.047	0.951	0.912	0.992	95.7	0.953	0.116	0.925	0.841	0.921	92.0
5×10^{-4}	0.294	0	0.455	0.434	0.636	68.7	0.783	0	0.878	0.797	0.895	88.8

Table 4 Performance analysis on different databases (fold-wise) at $\alpha = 7 \times 10^{-4}$

	Performance evaluation parameters											
	MIAS						DDSM					
Folds	Testing instances	TP	FP	TN	FN	AUC	Testing instances	TP	FP	TN	FN	AUC
fold 1	12	6	0	6	0	1	25	12	1	12	0	1
fold 2	12	4	0	7	1	0.995	25	13	0	12	0	1
fold 3	12	5	0	7	0	1	25	12	0	12	1	1
fold 4	12	5	0	7	0	1	25	13	0	12	0	1
fold 5	12	6	0	6	0	1	25	13	0	12	0	1
fold 6	11	4	0	6	1	0.985	25	12	0	12	1	0.992
fold 7	11	5	0	6	0	1	25	13	0	12	0	1
fold 8	11	5	0	6	0	1	25	13	0	12	0	1
fold 9	11	5	0	6	0	1	25	13	0	12	0	1
fold 10	11	4	0	7	0	1	25	13	0	12	0	1

We have compared the performance of the present classifier with two other classifiers such as SVM and K-NNs and is shown in Table 5. It may be observed that the AdaBoost-RF classifier performs better than others.

We have also compared ROC curves achieved by the AdaBoost-RF classifier with that of SVM and K-NN; these are presented in Fig. 3. The optimal AUC values are 0.9985 and 0.9992 obtained by the AdaBoost-RF classifier for MIAS and DDSM databases, respectively. Table 6 presents the comparative analysis of various performance measures of the present scheme with the existing schemes. It may be observed that the suggested scheme outperforms its competent ones.

Table 5 Performances comparison of different classifiers at $\alpha = 7 \times 10^{-4}$

Database	classifier	Performance measures			
		F-measure	MCC	AUC	ACC, %
MIAS	Adaboost-RF	0.980	0.965	0.998	**98.3**
	K-NN	0.771	0.696	0.803	83.5
	SVM	0.603	0.545	0.682	74.8
DDSM	AdaBoost-RF	0.988	0.976	0.999	**98.8**
	K-NN	0.883	0.804	0.895	89.2
	SVM	0.684	0.586	0.760	75.2

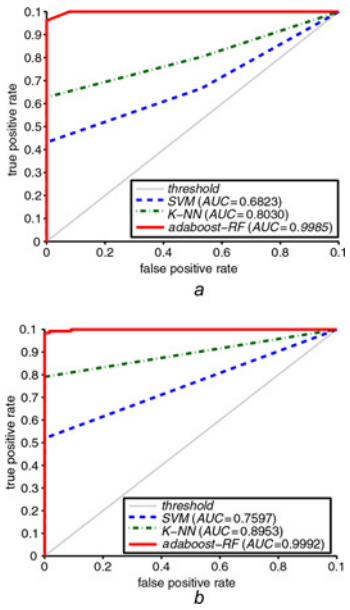

Figure 3 *ROC curves obtained by three classifiers at optimum significance level, $\alpha = 7 \times 10^{-4}$*
a MIAS database
b DDSM database

doi: 10.1049/htl.2014.0108

Table 6 Classification performance comparison between the proposed work and existing approaches

Approach	Technique	Database	Measurement
Verma *et al.* [5]	BI-RADS descriptor feature, SCBDL classifier	DDSM	ACC = 97.5%
Buciu and Gacsadi [7]	Gabor wavelets PCA and SVM	MIAS	AUC = 0.78
Xiaoming and Tang [9]	geometry and texture features SVM-RFE with NMIFS filter	DDSM	AUC = 0.9615
Görgel *et al.* [11]	SWT, SVM	MIAS, Istanbul University	ACC = 93.59%
Zanchetta *et al.* [10]	DWT, polynomial classifier	DDSM	AUC = 0.95%
Ganesan *et al.* [13]	trace transform, GMM	Singapore Anti-Tuberculosis Association CommHealth (SATA)	ACC = 92.48%
proposed work	DOST, null-hypothesis test with *t*-statistics AdaBoost-RF	MIAS, DDSM	ACC = 98.3%, AUC = 0.9985 (MIAS) ACC = 98.8%, AUC = 0.9992 (DDSM)

4. Conclusion: This Letter proposes an efficient scheme to classify mammographic images as benign or malignant to support breast cancer detection. The scheme utilises the DOST method to extract features from the mammographic images. A feature selection algorithm using the 'two-sample *t*-test' method is utilised to select the most discriminant features from the high-dimensional feature matrix. An AdaBoost algorithm is applied to classify the mammograms taking the random forest classifier as the base learner. The classification algorithm with the selected relevant features achieves the best performance at significance level, $\alpha = 7 \times 10^{-4}$. The optimal results achieved with respect to accuracy and AUC of ROC are 98.3% and 0.9985 for MIAS database. Similarly for DDSM database, the parameters are 98.8% and 0.9992, respectively. However, the extra feature selection is the additional overhead in the proposed scheme.

5. Funding and declaration of interests: Conflict of interest: none declared.

6 References

[1] Tabar L., Dean P.: 'Mammography and breast cancer: the new era', *Int. J. Gynecol. Obstet.*, 2003, **82**, (3), pp. 319–326
[2] Cheng H., Shi X., Min R., Hu L., Cai X., Du H.: 'Approaches for automated detection and classification of masses in mammograms', *Pattern Recognit.*, 2006, **39**, (4), pp. 646–668
[3] Liu S., Babbs C.F., Delp E.J.: 'Multiresolution detection of spiculated lesions in digital mammograms', *IEEE Trans. Image Process.*, 2001, **10**, (6), pp. 874–884
[4] Pereira R.R.Jr., Marques P.M.A., Honda M.O., *ET AL.*: 'Usefulness of texture analysis for computerized classification of breast lesions on mammograms', *J. Digit. Imaging*, 2007, **20**, (3), pp. 248–255
[5] Verma B., McLeod P., Klevansky A.: 'Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer', *Expert Syst. Appl.*, 2010, **37**, (4), pp. 3344–3351
[6] Fraschini M.: 'Mammographic masses classification: novel and simple signal analysis method', *Electron. Lett.*, 2011, **47**, (1), pp. 14–15
[7] Buciu I., Gacsadi A.: 'Directional features for automatic tumor classification of mammogram images', *Biomed. Signal Process. Control*, 2011, **6**, (4), pp. 370–378
[8] Prathibha B., Sadasivam V.: 'Hybrid transforms domain-based mammogram analysis using c-svm classifier', *Int. J. Med. Eng. Inf.*, 2012, **4**, (2), pp. 146–156
[9] Liu X., Tang J.: 'Classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method', *IEEE Syst. J.*, 2013, **8**, (3), pp. 1–11
[10] Zanchetto do Nascimento M., Martins A.S., Neves A.L., Ramos R.P., Flores E.L., Carrijo G.A.: 'Classification of masses in mammographic image using wavelet domain features and polynomial classifier', *Expert Syst. Appl.*, 2013, **40**, (15), pp. 6213–6221
[11] Görgel P., Sertbas A., Ucan O.N.: 'Mammographical mass detection and classification using local seed region growing-spherical wavelet transform (lsrg-swt) hybrid scheme', *Comput. Biol. Med.*, 2013, **43**, (6), pp. 765–774
[12] Kumar S.M., Balakrishnan G.: 'Multi resolution analysis for mass classification in digital mammogram using stochastic neighbor embedding'. Int. Conf. on Communications and Signal Processing (ICCSP), Melmaruvathur, India, April 2013, pp. 101–105
[13] Ganesan K., Acharya U.R., Chua C.K., Lim C.M., Abraham K.T.: 'One-class classification of mammograms using trace transform functionals', *IEEE Trans. Instrum. Meas.*, 2014, **63**, (2), pp. 304–311
[14] Drabycz S., Stockwell R.G., Mitchell J.R.: 'Image texture characterization using the discrete orthonormal S-transform', *J. Digital Imaging*, 2012, **22**, (6), pp. 696–708
[15] Groebner D.F., Shannon P.W., Fry P.C., Smith K.D.: 'Business statistics: a decision-making approach', 2005
[16] Breiman L.: 'Random forests', *Mach. Learn.*, 2001, **45**, (1), pp. 5–32
[17] Freund Y., Schapire R.E.: 'Experiments with a new boosting algorithm', *ICML*, 1996, **39**, pp. 148–156
[18] Friedman J., Hastie T., Tibshirani R.: 'Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors', *Ann. Stat.*, 2000, **28**, (2), pp. 337–407
[19] Duda R.O., Hart P.E., Stork D.G.: 'Pattern classification' (John Wiley & Sons, 2012)
[20] Suckling J., Parker J., Dance D., *ET AL.*: 'The mammographic image analysis society digital mammogram database', 1994
[21] Deserno T.M., Soiron M., de Oliveir J.E.: 'Texture patterns extracted from digitizes mammograms of different BI-RADS classes', Image Retrieval in Medical Applications project, release: V1.0, 2012, http://www.ganymed.imib.rwth-aachen.de/irma/datasets_en.php