International Journal of Innovative
Computing, Information and Control
Volume **13**, Number **3**, June **2017**

ICIC International ©2017 ISSN 1349-4198
pp. 967–978

EFFICIENT APPROACH FOR THE CLASSIFICATION OF MASSES IN DIGITAL MAMMOGRAMS

Shankar Thawkar[1] and Ranjana Ingolikar[2]

[1]Department of Information Technology
Hindustan College of Science and Technology
NH-2, Farah-Mathura, Uttar Pradesh 281122, India
Shankar.thawkar@gmail.com

[2]Department of Computer Science
St. Francis de Sales College
Seminary Hill, Nagpur, Maharashtra 440006, India
r_ingolikar@rediffmail.com

Received November 2016; revised March 2017

Abstract. *Breast cancer is still one of the leading diseases in women all over the world. Mammography is the best tool to discover breast tumor in its early stage. The computer based system will assist the radiologist in detection and classification of masses. These systems will help to improve breast cancer diagnosis and minimize unnecessary biopsies. In this work we investigate efficient methods for classification of masses into benign and malignant to improve breast cancer diagnosis. Initially twenty-five features based on intensity, texture and shape are extracted from each of the detected masses. Then six most significant features are selected by step-wise forward logistic regression technique. These features are used to train and test support vector machine, K-nearest neighbor and decision tree classifiers with 10-fold cross validation. The experiment was conducted on 651 mammogram images with 314 benign and 337 malignant cases obtained from digital database for screening mammography. The performance evaluation of classifiers indicates that SVM with radial bias function is better than both K-NN and decision tree classifier. Our method achieves best results of 97.32% sensitivity, 90.44% specificity and 94% accuracy for SVM with area under receiver operating characteristics curve* $A_Z = 0.963 \pm 0.008$. *All the results achieved are promising when compared with some existing works.*

Keywords: Mammograms, Mass classification, SVM, K-NN, DT, Logistic regression, Receiver operating characteristics curve

1. **Introduction.** Currently breast cancer is quite common compared to other cancers in women's. It appears to be one of the main causes of death of women worldwide [1-3]. At present no preventive method or technique is available that is why discovery of breast tumor in its premature stage plays a key role in effective diagnosis of breast cancer treatment. Mammography is the most excellent tool to discover breast tumor in its early stage. It enables to detect two most important symptoms of breast cancer such as masses and calcification [4]. The radiologists classify masses into benign and malignant by reading mammograms. However, reading of mammograms in early stage of breast cancer is a very challenging task for radiologist. The reading of mammograms to decide whether suspicious tissue is normal or malign depends on expertise and experience of radiologist. Even a specialist's inter observation rate varies [5]. Basically the tissue has been removed for the examination using biopsy technique. Statistic shows that more than 70% of biopsies of suspected breast cancer lesion turn out to be benign. A lot of research initiatives have been taken for the design and development of computer based system. These systems

assist radiologist in detection and classification of masses to minimize unnecessary biopsies and improve the breast cancer diagnosis. This research article investigates different classification techniques for achieving better performance of classifying masses into benign and malignant using optimal feature selection.

The remainder of the paper is organized as follows. Review of associated work done in this area is presented in Section 2. The proposed research methodology is presented in Section 3. Section 4 presents results of different methods and discussion about their performances and conclusion of the paper is shown in Section 5.

2. **Related Work.** Detection and classification of breast masses are the most challenging research area. Silva et al. [6] investigate the use of support vector machine (SVM) for the selection and classification of masses. The quality threshold is used to segment masses. The best mass candidates are selected by SVM. Haralick descriptors and a correlogram function are used to extract texture features from detected masses. Then SVM is used to classify masses using these features. The experiment was conducted on 599 images obtained from DDSM with 517 malign and 82 normal cases. The reported performance was 83.53% accuracy, 92.31% sensitivity, 82.2% specificity and area under receiver operating characteristics (ROC) curve 0.8003.

Petrosian et al. [7] presented a method for the classification of masses. A modified decision tree classifier was used for the classification of masses using eight textures based features. These features were calculated from SGLD matrix of each region of interest (ROI). The different combination of features is used to test performance of classifier. The classifier achieves 89% sensitivity and 76% specificity during training. With leave one out method the classifier performance during testing was about 76% sensitivity and 64% specificity. The experiment was conducted on 195 mammograms with 45 malign and 135 benign cases.

Martins et al. [8] proposed a method for the classification of masses that uses Ripley's K function and SVM. The suspicious region of mass is segmented with use of neural gas algorithm. The experiment was conducted on 997 images obtained from DDSM with 436 mammograms used for testing and 561 for training. The best result obtained with the proposed method was 89.3% sensitivity with 0.93 false positive and 0.02 false negative per image.

Nunes et al. [9] proposed a technique for selection and classification of masses using SVM. The suspected mass region was detected using K-mean algorithm and template matching. Texture and geometry based features are derived from detected masses. Then SVM is trained and tested using these features for the classification of masses into benign and malign. The experiment was conducted on 650 mammograms obtained from DDSM. The classifier achieves an accuracy of 83.94% with 83.24% sensitivity, 84.14% specificity, 0.55 false positive and 0.17 false negative per image. Martins et al. [10] investigate methods for detection and classification of masses. Mass segmentation is performed by K-mean algorithm. Then SVM is used to classify masses using shape and texture based features. The experiment was conducted on 1177 with 250 malign and 927 benign cases obtained from DDSM. The accuracy of classifier was 85%.

Ganesan et al. [11] presented a method for the classification of masses. Features based on higher-order spectra, local binary pattern and Law's texture energy are derived from detected masses. The rank based feature selection is carried out using several techniques such as sequential forward, backward and branch-and-bound. These optimal features act as input to six classifiers namely DT, fishers, LDA, SVM, nearest mean and Parzen. The experiment was conducted on two datasets, one obtained from DDSM with 300 masses and the other from Singapore Anti-Tuberculosis Association CommHealth database with 300

masses. The performance of classifiers was evaluated in terms of accuracy, sensitivity and specificity. Authors reported 91% accuracy of decision tree classifier on DDSM dataset and 96.8% on CommHealth database.

Lesniak et al. [12] investigated the benefit SVM for the classification of masses with 10-fold cross validation over a dataset of 1540 patients. The performance of SVM was compared with ANN, K-NN and LDA based on two sets of region-based features. The best mean exam sensitivities found are 0.545, 0.636, 0.648, and 0.675 for LDA, k-NN, ANN net and SVM respectively.

Chan et al. [13] studied the importance of texture features derived from GLCM matrix for classification of masses using linear discriminant analysis (LDA). The five optimal features out of eight features are selected by stepwise linear discriminant analysis. The experiment was conducted on 168 malign and 504 normal cases. The accuracy of the classifier was evaluated using area under ROC curve and the average value of A_Z is 0.84 during training and 0.82 during testing.

Zheng et al. [14] proposed hybrid K-SVM for the classification of masses into benign or malign. The features are obtained by K-means algorithm for benign and malignant tumors separately. Then generalized SVM is used for the classification with 10-fold cross validation and achieves an accuracy of 97.38% when tested on WDBC dataset of 32 mammograms.

Chou et al. [15] studied step-wise logistic regression method for the diagnosis of breast tumor in ultrasound (US) technique. An experiment was conducted on 111 US images. The tumor is segmented manually for obtaining contour features from radial length of the tumor boundaries. The accuracy of the system was evaluated using ROC curve. The area under curve A_Z is 0.97 ± 0.013. Authors conclude that the proposed method is effective for the diagnosis of breast tumor as it has high accuracy and high negative predictive value.

McLaren et al. [16] proposed an artificial neural network (ANN) and logistic regression methods for the classification of masses in magnetic resonance imaging (MRI) technique. The experiment was conducted on 28 benign and 43 malignant cases. The four optimal features are selected from eight morphological, ten gray-level co-occurrence matrixes and fourteen law of texture features. The performance of the methods is evaluated using area under ROC curve. The analysis shows that both the methods yield similar results.

The objectives behind the proposed method are as follows.

- Improve the accuracy of SVM, K-NN and DT based classification system.
- Select most relevant features that will minimize misclassification rate.
- Required to use large and balance data set (benign and malignant) because unbalanced data set may hamper the performance of classifiers.
- Evaluate performance of classifiers using statistical parameters sensitivity, specificity, accuracy and area under curve.
- The performance of the methods is tested with additional parameters like false positive rate and false negative rate.
- Design and develop CAD system that will assist radiologist and improve breast cancer diagnostic.

3. **Research Methodology.** The proposed methodology uses forward logistic regression technique for the selection of optimal features. These optimal features act as an input to three classifiers namely support vector machine (SVM), K-nearest neighbor (K-NN) and decision tree (DT) which are used for the classification of masses. The proposed method is shown in Figure 1.

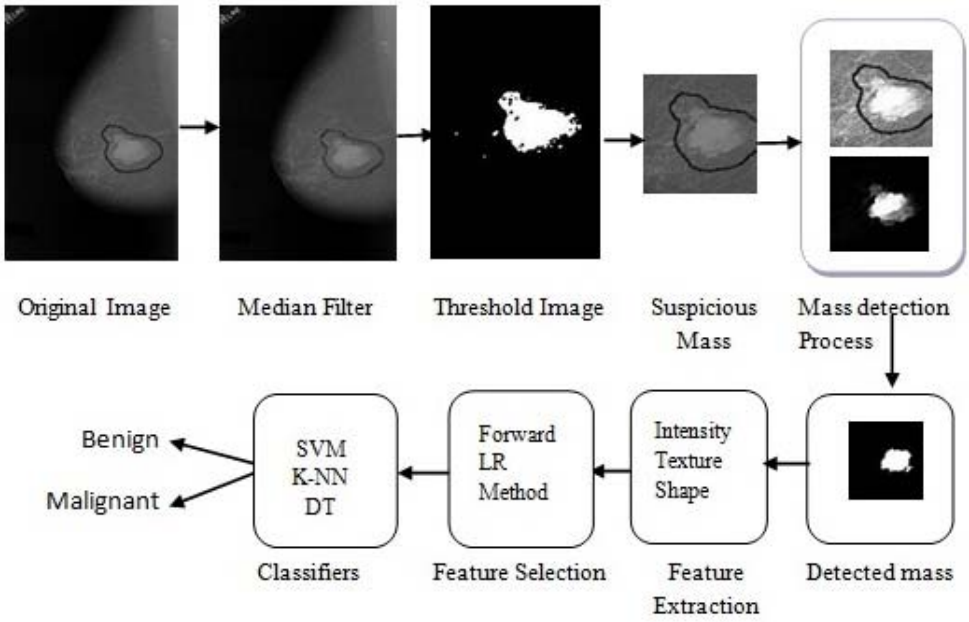


FIGURE 1. Overview of the proposed methodology

3.1. **Experimental database.** The proposed experiment was conducted on 651 mammograms obtained from digital database for screening mammography (DDSM). Out of 651 mammograms are 314 benign cases and 337 malignant cases. The database is available at www.marathon.csee.usf.edu/mammography/Database.htm. It consists of 2620 cases, classified into three cases normal, benign and malign [17]. The main objective of the database is to facilitate sound research in the development of algorithms for breast cancer diagnosis.

3.2. **Feature extraction.** The preprocessing of mammograms, mass segmentation and feature extraction is found in our previous work [18]. The features extracted are classified into three types: intensity features, textural features and shape features. Six intensity features are extracted from histogram analysis of detected masses. The textural based eleven features are extracted using gray level co-occurrence matrix (GLCM). The eight shape based features are extracted from shape of detected masses. The list of extracted features is shown in Table 1.

3.3. **Feature selection.** The enormous set of feature extracted from detected masses may degrade the performance of classifiers. The performance of the classifiers will be

TABLE 1. Extracted features

Type	Features
Intensity	Average gray level, Average contrast, Smoothness, Skewness, Uniformity, Entropy1 [4,21].
Texture	Energy, Entropy1, Contrast, Mean, Standard deviation, Variance, Correlation, Homogeneity, Sum average, Sum variance and Sum entropy [22,23].
Shape	Area, Perimeter, Compactness, Normalized standard deviation (Dnrl), Area ratio (RA), Contour roughness (NRV) and Overlapping ratio (Mshape) [15,24,25].

improved by selecting subset of optimal features [19]. The four criteria discrimination, reliability, independence and optimality are used to select most significant features [20].

In this paper step-wise forward logistic regression method is used to select subset of optimal features from the set of twenty five features. It is the most popular model used in medical research. It does not make any assumption about the distribution of predictor variables. Logistic regression predicts a categorical or response variable y from a set x_i variables, called independent or predictor variables. In forward logistic regression method entry and exit of variable in model one at a time are determined by the statistical score of the variable. The model is constructed by an iterative maximum likelihood procedure and the performance of the model depends on the maximum log likelihood value of each predictor variable. The smaller the values are, the better the model is. Whenever the variable is entered in the model, its significance is determined by log likelihood ratio. It is the difference between base model and current model as shown in (1).

$$\text{Deviation} = -2\ln\frac{\text{likelihood of the current model}}{\text{likelihood of the base model}} \tag{1}$$

In LR model the deviation in two models is represented by chi-square statistics with DF degree of freedom. Table 2 describes the intercept-only model. The Wald test describes the importance of variables in the model. It is based on chi-square distribution at 1 DF degree of freedom.

TABLE 2. Intercept-only model

	B	S.E.	Wald	df	Sig.	Exp (B)
Step 0 Constant	.076	.080	.914	1	.339	1.079

TABLE 3. Forward logistic regression model summary

Step	−2Log likelihood	Improvement			Model			Correct Class %	Variables Selected
		Chi^2	df	Sig.	Chi^2	df	Sig.		
1	794.288	78.16	1	.000	78.163	1	.000	63.0	R
2	730.920	63.36	1	.000	141.53	2	.000	68.9	Entropy1
3	603.505	127.41	1	.000	268.94	3	.000	78.1	Compactness
4	482.787	120.71	1	.000	389.66	4	.000	84.4	Sum Entropy
5	381.680	101.10	1	.000	490.77	5	.000	89.7	Area
6	352.131	29.54	1	.000	520.32	6	.000	92.1	Dnrl

The model summary for the selection of features is shown in Table 3. The goodness of the model is tested by three parameters log likelihood, chi-square values and correct classification rate. As shown in Table 3, the log likelihood value is decreasing after adding the new feature in model. Similarly there was a significant change in chi-square values with 1DF degree of freedom before and after adding the feature variables. The model also indicates that if the variable is added in the model, the rate of classification increases significantly.

3.4. **Classification.** Once the optimal subset of features is selected, classifiers are used to classify the masses. A support vector machine (SVM), K-nearest neighbor (K-NN) and decision tree (DT) classifiers are used to classify masses into benign or malignant.

3.4.1. *Support vector machine (SVM).* The basics of support vector machines (SVM) have been developed by Cortes and Vapnik for solving classification task [26]. The basic goal of SVM is to find an optimal hyperplane. The optimal hyperplane means to separate the data points with maximal margin. The data points which are close to the maximal margin hyperplane are called support vectors. The distance between the separating hyperplane and data points is called margin of the SVM classifier. Let us assume P is a training pattern of n sets with class labels, $P = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where x_i is a pattern and $\boldsymbol{y_i} \in \pm\mathbf{1}$ is a class of labels and S' is a dot product space and then a hyperplane in the space S' can be defined as shown in (2).

$$\{x \in S' | w.x + b = 0\}, \quad w \in S', \quad b \in R \tag{2}$$

where w is a weight vector normal to the line and b is a bias.

SVM with two most commonly used kernel functions, radial bias function (RBF) and sigmoid has been used in this article. The RBF kernel K on two samples x_1 and x_2 is defined as shown in (3).

$$K(x_1, x_2) = \exp^{-\gamma \|x_1 - x_2\|^2} \tag{3}$$

Similarly, sigmoid kernel is defined as shown in (4).

$$K(x_1, x_2) = \tanh(\gamma(x_1.x_2) + c) \tag{4}$$

where γ is a user defined parameter.

3.4.2. *K-nearest neighbor (K-NN).* K-NN is the most fundamental and simplest classification algorithm. It is a non-parametric method, which performs classification using nearest training sample in feature space [27]. Suppose X_i is the point in question (test), $X_i = x_1, x_2, \ldots, x_n$ and X_j is the training point then K-NN algorithm measures the distance $D(X_i, X_j)$ between point in question X_i and training sample X_j to classify the new object based on the majority of K-NN category of Y attributes of training sample. The proposed method uses K-NN with Euclidean and Mahalanobis distance metrics. The Euclidean distance between X_i and X_j is measured as shown in (5).

$$D(X_i, X_j) = \sqrt{\sum_{I=1}^{n} (x_i - x_j)^2} \tag{5}$$

Similarly, the Mahalanobis distance between point P_1 and distribution D_1 is defined as shown in (6).

$$Dist = \sqrt{(x - m)^T C^{-1} (x - m)} \tag{6}$$

where x is the observations, m is the mean values of observations and C^{-1} is the inverse covariance matrix. Both x and m are vectors.

The K-NN algorithm is summarized as:

Determine K number of nearest neighbors
Compute distance $D(X_i, X_j)$
Compute K-minimum distance neighbors
Gather Y values of nearest neighbors
Use majority of nearest neighbors to predict value of query point

3.4.3. *Decision tree (DT).* Decision tree (DT) is a classification technique which consists of three types of nodes: a root node, internal node and leaf or terminal node. Each node has at most one parent and two or more children [28]. A DT classifier predicts the target from several input variables. An input feature is represented by non-leaf node and target class by leaf node. A tree learns by splitting a set into subset based on some rules. The splitting process is performed recursively and it is based on division and conquers technique. The goodness of the split is measured by some parameters such as Gini index. The more detailed explanation about the construction of decision trees is found in [29,30]. DT is the most widely used technique for classification.

3.4.4. *Performance evaluation of classifiers.* The performance of the classifiers is measured using following parameters as shown in (7)-(14). All the values of these parameters are determined from confusion matrix.

Sensitivity (TPR): It defines the amount of positive cases (malign) correctly classified as positive (TP) among total positive cases.

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

Specificity (TNR): It defines the amount of negative cases (benign) correctly classified as negative (TN) out of total negative cases.

$$TNR = \frac{TN}{TN + FP} \tag{8}$$

Accuracy (ACC): It defines the total amount of true positive and true negative cases (malign and benign) correctly classified as TP and TN among total positive and negative cases.

$$ACC = \frac{\text{True Positive + True Negative}}{\text{True Positive + True Negative + False Positive + False Negative}} \tag{9}$$

Positive predictive value (PPV): It defines the proportion of TP result out of all true positive results. It is also called as precision.

$$PPV = \frac{TP}{TP + FP} \tag{10}$$

Negative predictive value (NPV): It defines the proportion of TN result out of all true negative results.

$$NPV = \frac{TN}{TN + FN} \tag{11}$$

False discovery rate (FDR): It defines the proportion of TN result out of all true negative results.

$$FDR = \frac{FP}{FP + TP} \tag{12}$$

False negative rate (FNR): It defines the proportion of TN result out of all true negative results.

$$FNR = \frac{FN}{TP + FN} \tag{13}$$

F_1 score: It is a measure of test accuracy.

$$F_1\ Score = \frac{2TP}{2TP + FP + FN} \tag{14}$$

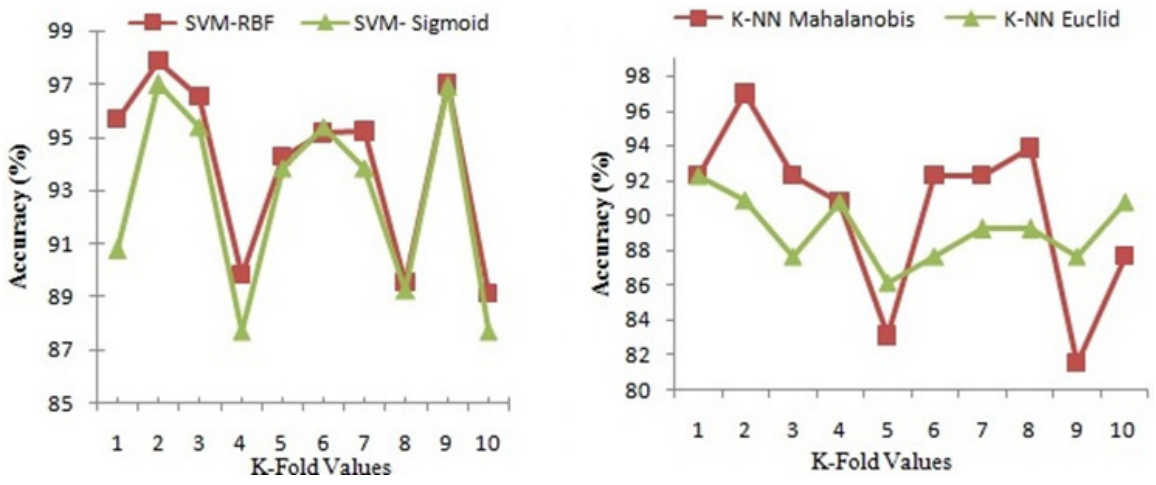
Area under ROC curve (AUC): It is another way besides confusion matrices to measure the accuracy of classifiers [31]. It is a plot between sensitivity (TPR) on Y axis and

1-Specificity on the X axis. The value of AUC lies between 0 and 1. If its value is 1, then model is 100% accurate.

4. **Results and Discussion.** The proposed experiment was conducted on 651 mammogram images with 314 benign and 337 malignant cases. As discussed in Section 3.2, total twenty five features based on intensity, texture & shape are computed from detected masses of 651 mammograms. The optimal features are selected with step-wise forward logistic regression method. As explained in Section 3.3, a subset of six optimal features is selected from a set of twenty five features. These six optimal features act as an input to three classifiers SVM, K-NN and DT, which are used to classify masses. The validity of the results produced by the classifiers is guaranteed by K-fold cross validation technique. In this method original data sample is randomly divided into K sub-samples. Of the K subsamples, one subsample is used to test (validate) the classifier and remaining $K-1$ samples are used to train the classifier. The validation is repeated K times and then all the K-folds results are averaged to find the accuracy of classifier.

In this article SVM is implemented with two kernel methods radial bias function (RBF) and sigmoid. The parameters Gama (γ), C are set to 0.05 and $C = -0.7$. All the parameters are determined using trial and error method. It has been observed that the best results are obtained with these parameters. The comparison of their performances for 10-fold cross validation is shown in Figure 2(a). The best classification accuracy of 97.87% is achieved for SVM-RBF at $K = 2$. The maximum number of miss-classifications occur at $K = 4$, 8 and 10 for both SVM-RBF and SVM-sigmoid. We can observe that SVM-sigmoid is better than SVM-RBF by 0.24% at $K = 6$. The average miss-classification rate for SVM-RBF and SVM-sigmoid is 6% and 7.2% respectively. Similarly average classification accuracy for SVM-RBF and SVM-sigmoid is 94% and 92.62% respectively. With the comparison of SVM-RBF and SVM-sigmoid, we conclude that SVM-RBF is better than SVM-sigmoid with 10-fold cross validation.

The K-nearest neighbor classifiers were implemented with two distance matrices Mahalanobis and Euclid. The parameter number of neighbors is set to 15 for predicting the samples. The performance comparison of both the algorithms for 10-fold cross validation is shown in Figure 2(b). One can observe that the accuracy of K-NN using Mahalanobis



(a) (b)

FIGURE 2. K-fold validation: (a) comparison of SVM-RBF and SVM-sigmoid for 10-fold validation; (b) comparison of K-NN Mahalanobis and K-NN Euclid for 10-fold validation

distance matrix is 92.31% at $K = 1$ and then rise by 4.66% at $K = 2$ (96.97%). The accuracy of classifier was oscillated largely up to $K = 9$ and reached minimum value of 81.54%. While in case of K-NN with Euclid distance matrix there is a steady change in accuracy as compared with K-NN Mahalanobis. The average miss-classification rate for K-NN Mahalanobis is 9.68% and for K-NN Euclid is 10.75%. Although the accuracy of K-NN Mahalanobis reached a minimum value of 83% and 81.54% at $K = 5$ and $K = 9$, it is better than K-NN Euclid.

The performance summary of three classifiers SVM with RBF kernel, K-NN with Mahalanobis distance matrix (MDM) and decision tree is presented in Table 4 as sensitivity, specificity, overall accuracy, PPV, NPV, false detection ratio, false negative ratio and F_1 score. One can observe from Table 4 SVM-RBF is better than all the classifiers with respect to measuring parameters. Our model SVM-RBF achieves accuracy of 94%, sensitivity of 97.32% and specificity of 90.44% with 91.62% of positive predictive values and 96.92% of negative predictive value. The model demonstrates lower FDR and FNR values with high F_1 score. These results indicate that model is the most suitable to implement as CAD and acts as a second reader for radiologist. The results also indicate that K-NN (MDM) is better than DT classifier. K-NN classifier achieves an accuracy of 90.32%. Although DT classifier shows poor performance as compared with SVM and K-NN, it is better when we compare it with results of related literature.

Another important parameter to express the accuracy of classifiers is area under ROC curve. The comparison plot for ROC curves of all the classifiers is shown in Figure 3

TABLE 4. Performance summary of classifiers

Method	TP	FN	TN	FP	Sen. TPR (%)	Spec. TNR (%)	ACC (%)	PPV	NPV	FDR	FNR	F_1 Score
SVM (RBF)	328	30	284	09	97.32	90.44	94.00	91.62	96.92	0.083	.028	94.38
K-NN (MDM)	321	47	267	16	95.25	85.03	90.32	87.22	94.34	0.127	.047	90.16
DT	286	51	263	51	84.86	83.75	84.33	84.86	83.75	0.151	.151	84.86

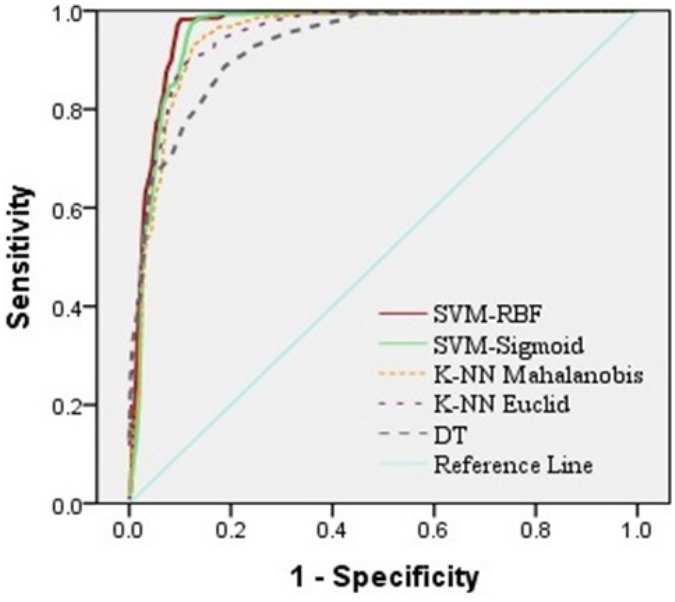


FIGURE 3. Comparison plot of ROC curves

and the calculated area under ROC curve (AUC) in Table 5. As indicated in Figure 3 and Table 5 AUC values for SVM-RBF, KNN-MDM and DT are $A_Z = 0.963 \pm 0.008$, 0.947 ± 0.009 and 0.929 ± 0.010.

Table 5 shows that K-NN with Euclid distance matrix is better than K-NN with Mahalanobis distance matrix. AUC value of K-NN Euclid is 0.951 ± 0.008. The highest AUC value achieved with SVM-RBF is $A_Z = 0.963 \pm 0.008$. As we can observer from AUC values, SVM-RBF is better than all the classifiers.

The comparison of the proposed method with related studies is presented in Table 6. It has been observed that different authors used different features with different database differed in number of benign and malignant cases. Even the methodology used for comparing performance of classifiers differs. According to [32,33] use of unbalanced cases (benign & malignant) or different database will affect the performance of classifiers. Most of the methodologies mentioned in Table 6 use accuracy as performance criteria, some of them

TABLE 5. Area under ROC curve

Test Result Variable (s)	Area	Std. Error	Sig.	Asymptotic 95% C.I.	
				LB	UB
SVM-RBF	.963	.008	.000	.947	.978
SVM-sigmoid	.955	.009	.000	.937	.972
K-NN Mahalanobis	.947	.009	.000	.930	.964
K-NN Euclid	.951	.008	.000	.935	.967
DT	.929	.010	.000	.910	.947

TABLE 6. Comparison of results

Author	Method	Data Set /features	No. of Cases Used			Sen. TPR (%)	Spec. TNR (%)	Acc. (%)	AUC
			B*	M*	Total				
Silva et al. [6]	SVM	DDSM/Texture	82	517	599	92.3	82.2	83.53	0.8003
Petrosian et al. [7]	DT	MIAS/Texture	45	135	180	–	89	76	–
Martins et al. [8]	SVM	DDSM	–	–	997	89.3	–	–	0.93FPR 0.02FNR
Nunes et al. [9]	SVM	DDSM	–	–	650	84.14	83.24	83.94	0.55FPR 0.17FNR
Martins et al. [10]	SVM	DDSM/Shape & Texture	927	250	1177	–	–	85	
Ganesan et al. [11]	DT	DDSM/Spectra, Local binary, Laws Texture energy	–	–	300	–	–	91	–
Lesniak et al. [12]	K-NN NN SVM	DDSM/Region Based	–	–	300	–	.636 .648 .675	–	–
Zheng et al. [14]	K-SVM	WDBC	–	–	32	–	–	97.38	–
Proposed Method	**SVM K-NN DT**	DDSM/Intensity, Texture & shape	314	337	651	97.32 95.25 84.86	90.44 85.03 83.75	94.0 90.32 84.33	0.963 0.947 0.929

B*: benign cases, M*: malignant cases

use sensitivity and specificity and very few use AUC as criteria for measuring performance of classifiers. One can observe from Table 6 our method SVM with RBF kernel achieves a highest accuracy 95% and AUC value $A_Z = 0.963 \pm 0.008$ which is better than all the methods cited in Table 6 when we compare them with parameters sensitivity, specificity, accuracy and AUC values. The performance of our method is a little bit poorer than the scheme investigated by Zheng et al. [14]. The cited method achieves an accuracy of 97.38% by K-SVM classifier when implemented on a dataset of 32 cases from WDBC.

If we compare the results with sensitivity of our method, then both the methods produce approximately the same results (97.32%). Although the results cited by [14] are good, our method is better than the cited method. The reason is that the cited method is implemented on only 32 cases and performance of classifier is measured as accuracy while our proposed method was implemented on 651 cases with sensitivity, specificity, accuracy and AUC value as performance criteria.

5. **Conclusion.** In this paper we have investigated an efficient method for classification of masses in digital mammograms. The six most significant features out of twenty-five features were selected using step-wise forward logistic regression technique. These six features are used to train and test three classifiers SVM, K-NN and DT with 10-fold cross validation. The outcomes of the experiment show that SVM with RBF kernel is better than K-NN and DT. SVM-RBF has the highest classification accuracy of 94% with AUC value $A_Z = 0.963 \pm 0.008$. All the results achieved are promising when compared with existing work. The proposed method helps to minimize unnecessary biopsies and improve the breast cancer diagnosis. In feature investigation we pay more attention to improving specificity (TNR) with the use of more efficient feature selection techniques.

REFERENCES

[1] N. C. I. (NCI), *Cancer Stat Fact Sheets: Cancer of the Breast*, http:// www.seer.cancer.gov/statfacts /html/breast.html, 2009.

[2] B. Acha, C. Serrano, R. M. Rangayyan and J. L. Desautels, Detection of microcalcifications in mammograms, *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*, Bellingham, 2006.

[3] R. L. Siegel, K. D. Miller and A. Jemal, Cancer statistics, 2015, *CA: A Cancer Journal for Clinicians*, vol.65, no.1, pp.5-29, 2015.

[4] H. D. Cheng, X. Cai, X. Chen, L. Hu and X. Lou, Computer-aided detection and classification of microcalcifications in mammograms: A survey, *Pattern Recognition*, vol.36, no.12, pp.2967-2991, 2003.

[5] P. Skaane, K. Engedal and A. Skjennald, Interobserver variation in the interpretation of breast imaging: Comparison of mammography, ultrasonography, and both combined in the interpretation of palpable non calcified breast masses, *Acta Radiologica*, vol.38, no.4, pp.497-502, 1997.

[6] J. Silva, A. O. C. Filho, A. C. Silva, A. C. Paiva and M. Gattass, Automatic detection of masses in mammograms using quality threshold clustering, correlogram function, and SVM, *Journal of Digital Imaging*, vol.28, no.3, pp.323-337, 2015.

[7] A. Petrosian, H. P. Chan, M. A. Helvie, M. M. Goodsitt and D. D. Adler, Computer-aided diagnosis in mammography: Classification of mass and normal tissue by texture analysis, *Physics in Medicine and Biology*, vol.39, no.12, p.2273, 1994.

[8] L. de Oliveira Martins, G. B. Junior, E. C. da Silva, A. C. Silva and A. C. de Paiva, Classification of breast tissues in mammogram images using Ripley's K function and support vector machine, *Image Analysis and Recognition*, pp.899-910, 2007.

[9] A. P. Nunes, A. C. Silva and A. C. Paiva, Detection of masses in mammographic images using simpson's diversity index in circular regions and SVM, *Machine Learning and Data Mining in Pattern Recognition*, pp.540-553, 2009.

[10] L. O. Martins, G. B. Junior, A. C. Silva, A. C. Paiva and M. Gattass, Detection of masses in digital mammograms using K-means and support vector machine, *ELCVIA: Electronic Letters on Computer Vision and Image Analysis*, vol.8, no.2, pp.39-50, 2009.

[11] K. Ganesan, R. U. Acharya, C. K. Chua, L. C. Min, B. Mathew and A. K. Thomas, Decision support system for breast cancer detection using mammograms, *Proc. of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 2013.

[12] J. Lesniak, R. Hupse, M. Kallenberg, M. Samulski, R. Blanc, N. Karssemeijer and G. Székely, Computer aided detection of breast masses in mammography using support vector machine classification, *SPIE Medical Imaging*, 2011.

[13] H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt and N. Petrick, Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space, *Physics in Medicine and Biology*, vol.40, no.5, p.857, 1995.

[14] B. Zheng, S. W. Yoon and S. S. Lam, Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms, *Expert Systems with Applications*, vol.41, no.4, pp.1476-1482, 2014.

[15] Y. H. Chou, C. M. Tiu, G. S. Hung, S. C. Wu, T. Y. Chang and H. K. Chiang, Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis, *Ultrasound in Medicine & Biology*, vol.27, no.11, pp.1493-1498, 2001.

[16] C. E. McLaren, W. P. Chen, K. Nie and M. Y. Su, Prediction of malignant breast lesions from MRI features: A comparison of artificial neural network and logistic regression techniques, *Academic Radiology*, vol.16, no.7, pp.842-851, 2009.

[17] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer Jr, R. Moore, K. Chang and S. Munishkumaran, Current status of the digital database for screening mammography, *Digital Mammography*, pp.457-460, 1998.

[18] S. Thawkar and R. Ingolikar, Automatic detection and classification of masses in digital mammograms, *International Journal of Intelligent Engineering and System*, vol.10, no.1, pp.65-74, 2017.

[19] M. Sameti, R. K. Ward, J. Morgan-Parkes and B. Palcic, A method for detection of malignant masses in digitized mammograms using a fuzzy segmentation algorithm, *Proc. of the 19th Annual International Conference of the IEEE*, vol.2, pp.513-516, 1997.

[20] H. Li, Y. Wang, K. J. Liu, S. C. B. Lo and M. T. Freedman, Computerized radiographic mass detection. II. Decision support by featured database visualization and modular neural networks, *IEEE Trans. Medical Imaging*, vol.20, no.4, pp.302-313, 2001.

[21] N. Petrick, H. P. Chan, B. Sahiner and M. A. Helvie, Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms, *Medical Physics*, vol.26, no.8, pp.1642-1654, 1999.

[22] R. M. Haralick, K. Shanmugam and I. H. Dinstein, Textural features for image classification, *IEEE Trans. Systems, Man and Cybernetics*, no.6, pp.610-621, 1973.

[23] R. M. Haralick, Statistical and structural approaches to texture, *Proc. of the IEEE*, vol.67, no.5, pp.786-804, 1979.

[24] A. V. Alvarenga, W. C. A. Pereira, A. F. C. Infantosi and C. M. Azevedo, Morphologic operators applied to breast tumour ultrasound image classification, *Acoustical Imaging*, pp.463-470, 2004.

[25] P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer Science & Business Media, 2013.

[26] C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning*, vol.20, no.3, pp.273-297, 1995.

[27] W. C. Lin, S. W. Ke and C. F. Tsai, CANN: An intrusion detection system based on combining cluster centers and nearest neighbors, *Knowledge-Based Systems*, vol.78, pp.13-21, 2015.

[28] M. A. Friedl and C. E. Brodley, Decision tree classification of land cover from remotely sensed data, *Remote Sensing of Environment*, vol.61, no.3, pp.399-409, 1997.

[29] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, John Wiley & Sons, 2014.

[30] A. E. G. Rodríguez, J. F. M. Trinidad, M. G. Borroto and J. A. C. Ochoa, Mining patterns for clustering on numerical datasets using unsupervised decision trees, *Knowledge-Based Systems*, vol.82, pp.70-79, 2015.

[31] J. A. Swets, Measuring the accuracy of diagnostic systems, *Science*, vol.240, pp.1285-1293, 1988.

[32] M. Kubat, R. C. Holte and S. Matwin, Machine learning for the detection of oil spills in satellite radar images, *Machine Learning*, vol.30, nos.2-3, pp.195-215, 1998.

[33] M. Kubat and S. Matwin, Addressing the curse of imbalanced training sets: One-sided selection, *ICML*, vol.97, pp.179-186, 1997.