

Characterization of Architectural Distortion in Mammograms Based on Texture Analysis Using Support Vector Machine Classifier with Clinical Evaluation

Amit Kamra¹ · V K Jain² · Sukhwinder Singh³ · Sunil Mittal⁴

Published online: 3 July 2015

© Society for Imaging Informatics in Medicine 2015

Abstract Architecture distortion (AD) is an important and early sign of breast cancer, but due to its subtlety, it is often missed on the screening mammograms. The objective of this study is to create a quantitative approach for texture

Highlights

1. Architectural distortion is the most difficulty abnormality to classify due to its varying attributes.
2. The system is trained and tested with images from three databases. Two databases viz. MIAS and DDSM are standard public domain databases. For clinical evaluation, the experiment has been done on actual patient database obtained from ACE Healthways, Diagnostic Centre Ludhiana, India.
3. The system has been tested with two types of ROI viz. fixed size and Ground (varying) truth.
4. Three features models have been used, GLCM, Fractal features and Fourier power spectrum.
5. A unique combination of texture based features has been proposed after reducing with Stepwise Regression based Feature Selection method.
6. The results have been validated with different performance evaluation parameters using support vector machine classifier using sequential minimal optimization.
7. The results have been validated statistically based on various statistical parameters.
8. Above all the accuracy of proposed classification system is quite high and comparable to work of other researchers
9. Results of clinical evaluation done on actual patient care signify that proposed study can be very useful in providing second opinion to the radiologists.

classification of AD based on various texture models, using support vector machine (SVM) classifier. The texture analysis has been done on the region of interest (ROI) selected from the original mammogram. A comprehensive analysis has been done on samples from three databases; out of which, two data sets are from the public domain, and the third data set is for clinical evaluation. The public domain databases are IRMA version of digital database for screening mammogram (DDSM) and Mammographic Image Analysis Society (MIAS). For clinical evaluation, the actual patient's database has been obtained from ACE Healthways, Diagnostic Centre Ludhiana, India. The significant finding of proposed study lies in appropriate selection of the size of ROIs. The experiments have been done on fixed size of ROIs as well as on the ground truth (variable size) ROIs. Best results pertain to an accuracy of 92.94 % obtained in case of DDSM database for fixed-size ROIs. In case of MIAS database, an accuracy of 95.34 % is achieved in AD versus non-AD (normal) cases for ground truth ROIs. Clinically, an accuracy of 88 % was achieved for ACE dataset. The results obtained in the present study are encouraging, as optimal result has been achieved for the proposed study in comparison with other related work in the same area.

✉ Sukhwinder Singh
sukhdalip@pu.ac.in
Amit Kamra
amit_kamra@gndec.ac.in
V K Jain
vkjain27@yahoo.com
Sunil Mittal
drsnilmittal@yahoo.com

¹ Department of Information Technology, Guru Nanak Dev Engineering College, Ludhiana, India

Keywords Architecture distortion · Texture features · Classification · Stepwise regression · Clinical evaluation

² Department of Electrical and Instrumentation Engineering, SLIET, Deemed University, Sangrur, India

³ Department of Computer Science and Engineering, UIET, Panjab University, Chandigarh, India

⁴ Department of Radiology, ACE Healthways Diagnostics Centre, Ludhiana, India

Introduction

Breast cancer is the second most common cancer after cervical cancer prevailing among women in India [1, 2]. Although the last decade has observed improvement in treatment of breast cancer, still the diagnostic performance of imaging modalities is objected by subtle findings, called architecture distortion (AD). AD is the third most common cause of false negatives on screening mammograms after masses and micro-calcifications. But due to its subtlety and varying attributes, it is often missed during screening. According to breast imaging reporting and data system (BI-RADS), architectural distortion is defined as follows: the normal architecture of breast tissue is distorted with no definite mass visible [3]. The presence of architectural distortion leads to a variation of the normal orientation field in the mammogram; therefore, the texture associated with this abnormality is also called oriented texture. The orientation field in any mammogram is described as a map that depicts the orientation angle of texture corresponding to each pixel [4].

Though a substantial record exists in the literature regarding the classification of masses and micro-calcifications, very few research endeavors have been reported for the characterization of architectural distortion in mammograms. Guo et al. [5] proposed an approach for the detection of AD in images done using Hausdorff dimension to define the texture feature of regions of interest (ROIs). A total of 40 ROIs of size 128×128 pixels were selected from MIAS database to evaluate the performance. Although an accuracy of 72 % was achieved, the approach was tested on smaller database. Nandi et al. [6] presented an approach for classifying mammograms using genetic programming with ability of feature extraction. In their method, the student *t* test was used for feature selection with divergence as separability measure. A total of 57 mammograms selected from the screen test Alberta are classified into normal and abnormal categories. Ayres et al. [7] proposed a CAD system using Gabor filter and phase portrait for detection of AD. Feature extraction was done using gray level co-occurrence matrix (GLCM) model with optimal feature selection done using stepwise regression to classify the ROIs of size 230×230 pixels from MIAS database. Support vector machine (SVM) was used as a classifier, and sensitivity of 84 % was obtained. Banik et al. [8] evaluated the concept of prior mammograms for the detection of architectural distortion in 106 prior mammograms obtained from the screen test Alberta. The authors extracted GLCM and fractal features from ROIs of size 128×128 pixels. Sensitivity of 0.8 at 5.7 FP/image was obtained. The methodology proposed by [7, 8] suffers from the limitation that the usage of phase portraits to represent the patterns of subtle signs does not tend to create a well-defined converging pattern. Buciu and Gacsadi [9] proposed an automatic approach to retrieve the directional features in mammograms from MIAS database, which were

further filtered by Gabor wavelets. The size of ROIs used for feature extraction was kept at 140×140 pixels. But no criterion was given in selecting the size of ROI. Sensitivity of 76.92 % was obtained. Elthoukhy et al. [10] presented an approach based on multi-scale method to transform the mammograms into coefficients. ROIs of size 128×128 pixels from MIAS dataset were used for testing and training. The texture feature selection was done through Fischer's discriminant ratio and student's *t*-test. An accuracy of 95.98 % was achieved using five cross validation.

Previous studies have shown that the characterization of architectural distortion is a tedious task due to its inherent subtlety and varying attributes [3, 7, 8]. Clinically, radiologists often identify textural information in the form of qualitative traits (spiculations radiating from a point, focal distortion at the edge of parenchyma), which is highly subjective. The selection of the appropriate size of ROI remains to be a hot topic of research. Considering the subtle nature of AD, it is difficult to choose the reliable features and accordingly identify the optimal features that are able to classify the abnormality on the basis of texture features.

In this study, the feature extraction, selection, and classification have been done on two types of ROIs. From the literature survey, it was found that very few researchers have considered both the ground truth-size-based ROIs and fixed-sized ROIs. Ground truth refers to an actual size and location of the ROI. Fixed size refers to some predefined size of ROI which includes abnormality as well as some surrounding region in the proximity of abnormality. Unlike fixed-size ROIs, ground truth ROIs selected are of variable size but are extracted in square form to retrieve the texture features. Thus, the current study proposes multi-size ROI analysis to investigate texture properties of the tissue surrounded by subtle signs. In addition, the validation of current study has been clinically evaluated on the actual patient database.

This paper analyzes the problem of classifying the mammographic ROI into two classes, i.e., AD (abnormal) vs. non-AD (normal). Visually, both the benign and malignant ROIs appear to be more or less the same; therefore, more accurate decision to differentiate between benign and malignant tissues is left to the discretion of domain expertise. Thus, the main objective in the present study is to apply the CAD procedures to assist in the proper interpretation of AD and non-AD breast tissues. An extensive set of textural feature extractors has been applied on the ROIs. Various feature extraction models used in the proposed study are spatial gray level co-occurrence matrix (SGLCM) [11], Fourier power spectrum [12], and fractal features [13, 14]. Stepwise regression was applied to retrieve the most prominent features [15, 16]. Finally, the optimal texture features were applied to linear and non-linear kernels of the SVM classifier to evaluate the objective performance of each dataset [17]. The results are obtained by fine-tuning the SVM through various kernel parameters.

Materials and Methods

Image Dataset Used

The current study has been done using three data sets viz. image retrieval in medical application (IRMA) version of digital database for screening mammograms (DDSM) [18], Mammogram Image Analysis Society (MIAS) [19], and ACE dataset which is used for clinical evaluation that comprises of an actual patient database available from ACE Healthways, Diagnostics Centre Ludhiana, India. Unlike MIAS images that have a fixed resolution of 200 μm , DDSM images are scanned at different resolutions; Lumisys scanner has a 50- μm resolution and Howtek has 43.5 μm . The mammograms from the ACE dataset are digitized through Lorad M-IV series developed by Hologic USA. For the former two databases, the mammograms have ground truth provided by the experienced radiologists that includes the location of the abnormality, its radius, and type of abnormality. While for ACE dataset, the gold standard for the mammograms was created with the assistance of three expert radiologists. Architectural distortion being a subtle abnormality, therefore the ratings of subtlety were retrieved from all datasets. The proportion of subtle cases for DDSM and MIAS datasets is 91 and 68 %, respectively. While for ACE dataset, the fraction of samples corresponding to subtle cases of AD was 86 %.

Proposed Method

The main objectives of this study are as follows: (1) to differentiate AD ROIs from non-AD ROIs by means of feature extraction. (2) To study the effect of having fixed-size ROI and variable-size ROI on the classification accuracy. (3) To judiciously use feature selection that may improve the functioning and diminish the trickiness of the classifier. (4) To analyze the performance of selected features through various SVM kernels. (5) To clinically validate the results.

The flow chart of the proposed methodology is shown in Fig. 1 starting with ROI extraction, texture feature extraction, feature selection, and classification.

ROI Extraction

Unlike the masses, the area associated with architectural distortion does not have clear boundaries. Generally, the size of such region is inconsistent and ambiguous. The pathological area may vary, and size of tumor can be larger or smaller than normal region. This fact makes the size of ROI very significant, as it encloses the majority of abnormal tissues. To address this issue, the effect of ROI size on the texture features for evaluating the risk of breast cancer has been analyzed in the proposed study. The image-processing methods described in

our earlier work [20] including the application of Gabor filter and gradients were applied to mammogram samples from DDSM and MIAS datasets. The technique was based on the mapping of pixel orientations, wherein, if there is a decrease in the number of pixels pointing to a given region is found then that region is considered to be suspicious. The detected ROIs comprising of architectural distortion were labeled as true positive (TP) ROIs while the others were labeled as false positive (FP) ROIs. The ROIs detected by our method were of variable sizes. But in order to have a multi-size analysis, fixed size of ROIs was also cropped from original mammogram.

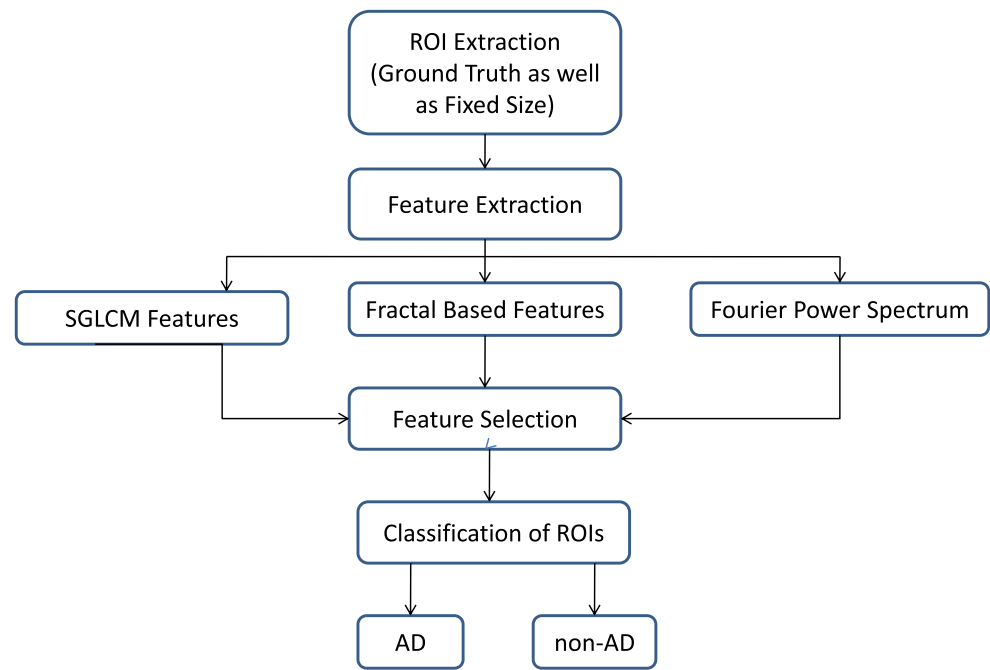
For DDSM data set, all the mammograms are of very large size varying from 3000 to 6000 pixels. Thus, there is a lot of redundant information. By examining the size of the abnormal region in about 147 samples showing architectural distortion and from the previous surveys [21, 22], an optimal size of 512×512 pixels was selected. To have a relative comparison with non-AD ROIs, the size of normal ROIs is also selected as 512×512 pixels. For the ground truth AD ROIs, the actual size was retrieved by the Gabor filter and gradient-based approach, but size of non-AD cases was kept at 512×512 pixels.

In case of MIAS database, it was found that some researchers has used ROIs of size 64×64 pixels, while some have done their analysis on 128×128 or 140×140 pixels [21, 22]. The main limitation in this database is that the actual number of samples depicting the pure architectural distortion is only 19. By analyzing the entire samples, it was found that the minimum and maximum value of radius encircling the abnormality is 23 and 117 pixels, respectively. The average radius of abnormality was found to be 32.5 pixels. Based on this, ROIs of size 32×32 pixels proved to be an optimum choice, as it ought to cover most of abnormal regions. In addition, this also gave the choice of having more samples for training and testing purpose by dividing larger ROIs into smaller ones. While for ground truth, the actual ROIs detected by our previous method is used [20]. This enables us to accommodate the maximum area comprising of the abnormality. The size of a normal (non-AD) ROI was selected as 64×64 pixels to have a relative comparison with the ground truth AD ROIs. The cropping of normal ROIs was done from the center of the breast, region behind the nipple, which comprises of ducts, ligaments, lymph nodes, usual parenchymal patterns, etc.

Texture Feature Extraction

In this phase, both the ground truth and fixed-size ROIs are processed separately for texture analysis. The texture analysis of ROI lies on the principle that if structure of breast tissue

Fig. 1 Block diagram of proposed classification system



reflects an altered mammogram, it will in turn give texture feature values different to normal tissue.

- (i) *SGLCM model*: as reported during literature survey, many researchers have used GLCM model for feature extraction. In the proposed study, all the 14 features retrieved from the co-occurrence matrix are estimated, but here, the pixel distances are varied with $D=1, 2, 3, 4, 5$ quantized into four directions ($\theta=0, 45, 90, 135^\circ$) [23]. Thus, for each ROI, 280 texture features are extracted. The reason for choosing this variation in distance is due to the fact that a single displacement value is not sufficient to elaborate the type of texture that has been investigated.
- (ii) *Fractal-based features*: due to the presence of architectural distortion, there is a disruption in normal tissue patterns. Fractal-based features can be significant in distinguishing between normal and tissue showing subtle signs [24, 25]. Two fractal-based features viz. Hurst coefficients at two resolutions were computed for both normal and abnormal ROIs.
- (iii) *Fourier power spectrum*: very few researchers have used the Fourier power spectrum features that are significantly correlated to the indications shown by AD, which includes changes in orientation of texture depicted by two features viz. radial angle and angular sum.

Prior to feature selection, all the features are normalized so as to have zero mean and unit variance [26].

Feature Selection

Out of 284 features extracted through various texture models, many features may turn out to be irrelevant as there was no significant difference between them. Initially, the experiments were done using Fischer's discriminant ratio (FDR) and sequential forward floating search (SFFS) method, but the outcome of these methods did not turn out optimal. Therefore, the modeling of texture to select appropriate attributes that can represent the subtle signs has been done through stepwise regression [7, 8]. Stepwise regression is a process of building a regression representation where the model is created by a group of candidate's predictor variable which are entered and removed from the model in a stepwise manner [27]. In this process, the features are added or eliminated from the model depending on their statistical significance in a regression. The criterion for adding/, removing features is based on the concept of F-statistics. The main advantage of stepwise regression includes regression models in which the choice of predictive variables is carried out by an automatic procedure. Other advantages include easily extended to other regression problems, works pretty well, particularly for large n , and can be improved by certain stopping rules.

Classification

The purpose of this phase is to classify the ROIs into AD and non-AD categories. During literature survey, it has been found that several researchers have used various classifiers for categorization of AD such as Fisher's linear discriminant analysis,

Bayesian classifier, and neural networks. But usually, the outcome of a classifier for non-diagnosed cases depends on the sample size used for training. The viability of samples is sometimes limited to the number of images available in the database. The same limitation was observed in the proposed study due to non-availability of large number of specific samples showing architectural distortion (especially in MIAS dataset). It has been found that SVM was able to provide generalized result, especially in the cases where numbers of samples were quite low [5]. Therefore, in the present study, SVM-based classifier with sequential minimal optimization (SMO) algorithm has been used to get accurate rate of classification. The main advantage of SMO is that it quickly solves the quadratic problem with no extra matrix storage and without invoking an iterative numerical routine for each sub-problem. To minimize the problem of not having a simple hyper-plane, an optimal setting of the various parameters for classifier has been done empirically using various kernels. The proper selection of kernel is significant as it specifies the adaptability of the final outcome of SVM in fitting the data [28].

Performance Evaluation

The performance evaluation is done to separate the positives from negatives. In this study, the abnormal ROIs are labeled as positives while the normal ROIs as negatives. The evaluation is done through six parameters viz. sensitivity, specificity, accuracy, F-score, Youden's index, and receiver operating characteristic (ROC) curve [29].

Sensitivity and specificity should be nearer to 1 for an excellent classification system. F-score is another performance measure for accuracy of a test. It is a weighted mean of recall and precision where the former refers to the number of right outcomes divided by the number of outcomes that should have been returned and the latter refers to the number of right outcomes divided by the number of all returned outcomes. The value of F-score approaching 1 is considered as the best value. Youden's index is a measure of overall diagnostic effectiveness. It ranges between 0 and 1, with values close to 1 indicates that the efficacy is relatively large and values close to 0 indicating limited effectiveness. ROC curve illustrates the sensitivity versus false positive rate (FPR) and is a useful analysis tool for binary classification problems.

Results and Discussions

It is to be noted that different features are selected depending upon the size, the distinctiveness of the datasets, and number of samples used in the training set. This uniqueness plays a significant role in pattern classification. In the present study,

all the datasets are divided into two parts, i.e., two-thirds of data are selected for training and the remaining one-third for testing the system [30]. The performance of all classification has been validated through threefold cross validation. The reason for choosing threefold is due to the fact that the number of experiments and consequently computation time are reduced. In addition, the variance of the estimator will be small, and the bias of the estimator will be large. As reported in literature, for large datasets, threefold cross validation is quite accurate [31]. The selection of samples belonging to testing and training is done randomly to rationalize any sort of partiality. The final results are based on the average results obtained from the test dataset for each fold. The detailed prediction results for the best outcome are also presented in the form of confusion matrices.

Classification Performance for MIAS Database

Fixed-Size ROIs (32×32 pixels)

Initially, the performance evaluation has been done on fixed-size ROIs from MIAS database. Optimal feature selection using stepwise regression leads to the selection of the following four features viz. maximum correlation coefficient ($\theta = 135^\circ$, $D=4$), inverse difference moment ($\theta = 90^\circ$, $D=5$), radial sum (S_r), and Hurst coefficient at first resolution (H_1).

Maximum correlation coefficient is a measure of gray level dependence between the pixels at the specific positions relative to each other. *Inverse difference moment* is a measure that reflects homogeneity of the ROI. Due to the presence of architectural distortion, the region becomes inhomogeneous; the region of distortion generally shows low IDM value and higher value for non distorted regions. Another feature *radial sum* is significant for subtle signs as it is able to measure the periodicity in a specific direction. *Hurst coefficient* is correlated to fractal dimension of the ROI. It has been reported in literature survey that fractal dimension of normal ROI varies significantly in comparison to the abnormal ROI. The larger the value of Hurst coefficient, the smaller is the fractal dimension, the smoother is the surface and vice versa. It is a significant attribute in discriminating the normal and abnormal breast parenchyma.

Ground Truth ROIs

In this phase, the analysis is done using ROI of actual size i.e., ground truth from MIAS database. The non-AD cases chosen are of size 64×64 pixels. Optimal features selected through stepwise regression comprises of the following three features viz. correlation ($\theta = 90^\circ$, $D=2$), Hurst coefficient at first resolution (H_1) and radial sum (S_r).

Discussion

Both the fixed and ground truth ROIs from MIAS database are processed separately. Table 1 shows the combined results for the various performance evaluation parameters for both fixed and ground truth ROIs.

(a) Fixed-Size ROIs For classification, threefold cross validation has been used. The results shown in Table 1 corresponding to fixed-size ROIs are the average of various results obtained through different combinations. The performance of an SVM depends largely on a number of parameters; therefore, various combinations of kernel parameters and penalty parameter “C” has been tried on the basis of experimentation. Best results are found with polynomial kernel with order 1. It is clearly visible from the table above that for linear kernel, all the abnormal cases were not classified correctly, due to subtle manifestations of TP ROIs. The other reason is that some of TP ROIs may perhaps comprise only a part or a small fraction of spicules arising from the focal point of AD; such ROIs add to the ambiguity and create intricacy in classification. These factors could be the underlying cause of the low value of the F-score. But even with increased ambiguity and the inclusion of normal cases, the rates of sensitivity have been obtained at nearly 90 % over the entire data sets. The results are further improved for radial basis function (RBF) and polynomial kernels where sensitivity of 92.30 % is obtained. It is inferred that the value of specificity obtained is relatively low. Specificity can be further improved at the later stage through biopsy. The ROC value near to 0.90 in case of RBF and polynomial kernels indicates a substantial reduction of the number of false positives per image in the detection of AD.

(b) Ground Truth ROIs Results shown in Table 1 is the average of all combinations obtained through threefold cross validation. It is evident from Table 1 that the number of

normal cases is quite higher than the abnormal ones, but this inclusion of higher number of normal cases did not substantially lower the performance of final outcomes. The results are quite promising for almost all the three kernels with best results pertaining to both RBF and polynomial kernel (order 1) with accuracy more than 95 %. From the experimental results, it is concluded that selected features corresponding to variable-sized ROIs perform very well, especially in classifying the mammograms into AD and non-AD. The main reason for this is that the samples from MIAS dataset might contain the abnormality concentrated in the center.

Classification Performance for DDSM Database

Fixed-Size ROIs (512 × 512 pixels)

In this phase, the experiment for feature extraction was performed on 221 fixed-size ROIs from DDSM dataset. On applying the feature selection using stepwise regression, the following set of optimal features were selected viz. sum variance ($\theta=45^\circ$, $D=5$), difference entropy ($\theta=45^\circ$, $D=5$), inverse difference moment ($\theta=135^\circ$, $D=5$), and Hurst coefficient at first resolution (H_1).

In comparison to MIAS dataset, the feature selection for DDSM dataset leads to the selection of two additional features viz. sum variance and difference entropy. *Sum variance* is able to characterize the subtle signs due to the fact that it imparts more stress on the elements that differs from the average value. During AD, the region where there is a higher concentration of distortion in oriented texture, sum variance is more predominant. *Difference entropy* represents the spread out of directional components in a ROI. On the other hand, if the ROI is composed of directional components with a uniform distribution, the entropy value will be higher. If the ROI is composed of directional components oriented in a very narrow angle band, the entropy value will be small.

Table 1 Classification performance for Fixed- and Ground Truth-Size ROIs from MIAS Database

Sr. no.	Size of ROI	AD	Non-AD	Kernel	Confusion matrix	Sensitivity (%)	Specificity (%)	Accuracy (%)	A_z	F-score	Y index
1	Fixed size	13	36	Linear	11(TP) 02(FN) 20(TN)	84.61	55.55	63.26	0.77	0.55	0.40
2	Fixed size	13	36	RBF	12(TP) 01(FN) 22(TN)	92.30	61.11	69.38	0.91	0.61	0.53
3	Fixed size	13	36	Polynomial	12(TP) 01(FN) 25(TN)	92.30	69.44	75.51	0.87	0.67	0.61
4	Ground Truth	07	36	Linear	05(TP) 02(FN) 35(TN)	71.42	97.22	93.02	0.95	0.76	0.68
5	Ground truth	07	36	RBF	06(TP) 01(FN) 35(TN)	85.71	97.22	95.34	0.98	0.85	0.82
6	Ground truth	07	36	Polynomial	06(TP) 01(FN) 35(TN)	85.71	97.22	95.34	0.98	0.85	0.82

Ground Truth ROIs

In this case, ROIs based on actual ground are selected for classification. But when the feature selection was applied to these ROIs, the features selected were same as that of fixed-size ROIs. The probable reason for same features being selected is that the average size of ground truth ROIs is approximately 512×512 pixels.

Discussion

Both the fixed and ground truth ROIs from MIAS database are processed separately. Table 2 shows the combined results for the various performance evaluation parameters for both fixed and ground truth ROIs.

(a) Fixed-Size ROIs Results in Table 2 shows that very promising results are obtained with all kernels providing with highest sensitivity of 96.66 % for RBF kernel with value of γ equal to 1 and penalty parameter “C” chosen as 0.8. It is evident from Table 2 that the numbers of TP ROIs are higher as compared to those for other subsets. Though, some cases of false negatives are detected, but it is due to the fact that few images showing architectural distortion contain more than one site of abnormal regions. In case of linear kernel, sensitivity of 91.66 % is quite good, but specificity obtained is rather low. The results are improved for RBF and the polynomial kernel where sensitivity is quite promising but specificity is slightly lower.

(b) Ground Truth ROIs It is evident from Table 2 that results are improved for RBF and polynomial kernels where accuracy of more than 85 % seems to be optimal in real-time scenario. Best results are achieved with RBF kernel with γ of 1.5 and penalty parameter “C” chosen as 0.5. The samples that have led to misclassification for FN may be having more than one set of AD or some scattered/subtle parts of AD patterns which

were not in correlation with the extracted features. In the case of linear kernel, results of sensitivity were not very good, though specificity achieved was excellent.

The multi-size analysis done on two standard datasets gives an idea that the abnormality could be either concentrated in the center of ROI or it might be scattered in whole ROI. Contrary to MIAS database, the results for DDSM samples in case of fixed-size ROIs are better. It is mainly due to the fact that the abnormality has been retrieved from the center of ROI. The results are slightly reduced in DDSM datasets for ground truth ROIs as some extraneous regions might have been added to the boundaries of extracted region. This extraneous region has lead to an increase of FP in the final classification process.

Statistical Significance

Basically, the statistical significance of any research problem is used to estimate the probability that any outcome observed in the data occurred only by chance. To nullify the effect of results being achieved by random chance, three cross validation has been used in the present work. The whole process has been repeated five times in which samples were randomly selected every time. This process itself avoids over fitting. The random samples have been used to mitigate any bias caused by the individual samples being chosen. The accuracy on the different iterations is averaged out to yield the overall accuracy. Thus, the final outcome is not just a chance occurrence.

To validate the results statistically, various statistical parameters like AUC (area under curve), 95 % confidence interval (CI), standard error (SE), and P values are being computed. A perfect test (one that has zero false positives and zero false negatives) has an AUC of 1.00. The 95 % CI is the interval in which the true (population) area under the ROC curve lies with 95 % confidence. The significance level or P value tests the null hypothesis that the AUC really equals 0.5.

Table 2 Classification performance for Fixed- and Ground Truth-Size ROIs from DDSM database

Sr. no.	Size of ROI	AD	Non-AD	Kernel	Confusion matrix		Sensitivity (%)	Specificity (%)	Accuracy (%)	A_z	F-score	Y index
1	Fixed size	60	25	Linear	55(TP) 05(FN)	05(FP) 20(TN)	91.66	80	88.23	0.91	0.91	0.71
2	Fixed size	60	25	RBF	58(TP) 02(FN)	06(FP) 19(TN)	96.66	76	90.58	0.95	0.93	0.72
3	Fixed size	60	25	Polynomial	56(TP) 04(FN)	02(FP) 23(TN)	93.33	92	92.94	0.95	0.94	0.85
4	Ground truth	60	25	Linear	34(TP) 26(FN)	0(FP) 25(TN)	56.66	100	69.4	0.88	0.72	0.56
5	Ground truth	60	25	RBF	53(TP) 07(FN)	05(FP) 20(TN)	88.33	80	85.88	0.89	0.89	0.68
6	Ground truth	60	25	Polynomial	52(TP) 08(FN)	03(FP) 22(TN)	86.66	88	87.05	0.91	0.90	0.74

Table 3 Analysis of statistical significance using the A_z , CI, SE, and P values

Sr. no.	Database	Size of ROI	Kernel	AUC	95 % CI	SE	P values (4 decimal places)	Comments
1	DDSM	Fixed size	Linear	0.91	0.86 to 0.97	0.029	0.0000	Excellent test
2	DDSM	Fixed size	RBF	0.95	0.91 to 0.99	0.020	0.0000	Excellent test
3	DDSM	Fixed size	Polynomial	0.95	0.91 to 0.99	0.020	0.0000	Excellent test
4	DDSM	Ground truth	Linear	0.88	0.81 to 0.96	0.039	0.0000	Good test
5	DDSM	Ground truth	RBF	0.89	0.82 to 0.96	0.033	0.0000	Good test
6	DDSM	Ground truth	Polynomial	0.91	0.84 to 0.97	0.030	0.0000	Excellent test
7	MIAS	Fixed size	Linear	0.77	0.55 to 0.84	0.074	0.0034	Fair test
8	MIAS	Fixed size	RBF	0.91	0.82 to 0.99	0.042	0.0000	Excellent test
9	MIAS	Fixed size	Polynomial	0.87	0.77 to 0.97	0.052	0.0036	Good test
10	MIAS	Ground truth	Linear	0.95	0.84 to 1.00	0.580	0.0000	Excellent test
11	MIAS	Ground truth	RBF	0.95	0.84 to 1.00	0.580	0.0000	Excellent test
12	MIAS	Ground truth	Polynomial	0.98	0.90 to 1.00	0.036	0.0000	Excellent test

If the P value is small, it may be concluded that the test actually does discriminate between AD and non-AD classes. Table 3 illustrates the statistical parameters computed for all the cases corresponding to standard datasets viz. MIAS and DDSM. As reported from Table 3, it is evident that the experimentation done in present work does have an ability to distinguish between the AD and non-AD classes. In almost all the cases, the AUC is more than 0.90; P values are smaller than 0.5, which shows that the result obtained has good discrimination between the two classes.

Clinical Evaluation

To further validate the results, clinical evaluation has been done to examine the efficacy of computer-based approaches to assess the presence or absence of any abnormalities in the actual patient database. All the clinically diagnostic images of mammograms were obtained from the radiology unit of ACE Healthways, Ludhiana, India. The necessary ethical permission for doing the research was obtained from the authorities of the imaging center before studying onset. In comparison to the public domain databases viz. DDSM and MIAS, there is no ground truth available in case of data sets from ACE Healthways. Therefore, to create the gold standard of the

image assessment of architectural distortion, three radiologists who have expertise in screening mammograms assessed all the mammograms and determined the presence of architectural distortion in them. Therefore, gold standard is generated manually, where the three radiologists were asked to classify the ROIs used for testing data set. To create a gold standard, an ROI was marked normal or abnormal, if it was judged similar by at least two radiologists. Further, to evaluate the ground truth (marked by the two radiologists), Gabor filter and gradient-based method has been used for those samples [20]. An audit of exposures to 43 women over the age of 35–60 years undergoing mammograms was carried out. All the mammograms were screened in the period from 1st June to 31st July 2014. The classification of mammograms was done on the data set of 73 ROIs (50 non-AD cases and 23 AD cases) from 43 mammograms of size 64×64 pixels. To have a comparative analysis, the normal ROI were kept of the same sizes. For classification, threefold cross validation has been used. Due to the lower number of AD cases, only fixed size of ROI has been selected for analysis. Table 4 gives the detail of the performance evaluation corresponding to ROIs for mammograms from ACE dataset.

Table 4 reveals that most of abnormal cases in ACE datasets are identified correctly. A sensitivity of 87.50 % was achieved with just one case of FNs. The results achieved

Table 4 Classification Performance for ROIs from ACE Dataset

Sr. no.	Size of ROI	AD	Non-AD	Kernel	Confusion matrix	Sensitivity (%)	Specificity (%)	Accuracy (%)	A_z	F-score	Y index
1	Fixed size	08	17	Linear	07(TP) 01(FN) 02(FP) 15(TN)	87.50	88.23	88	0.93	0.82	0.75
2	Fixed size	08	17	RBF	07(TP) 01(FN) 04(FP) 13(TN)	87.50	76.47	80	0.86	0.73	0.67
3	Fixed size	08	17	Polynomial	07(TP) 01(FN) 02(FP) 15(TN)	87.50	88.23	88	0.93	0.82	0.75

are quite promising in terms of ROC A_z value of 0.93 using the features selected by logistic regression with polynomial and linear kernel. The value of specificity can be improved through biopsy or by adding more representative samples in the testing set. The experts has defined 17 normal cases and eight abnormal cases (test cases) while the proposed system for classification using SVM classifier has classified 15 normal cases and seven abnormal cases correctly. The accuracy of technique was evaluated by comparison with its corresponding gold standard. An accuracy of 88 % is achieved, which is quite optimal, considering the subtle nature of the abnormality.

Comparison with Previous Works

The outcome of classification performance for various datasets pertaining to fixed-size and ground truth-based ROIs shows that the proposed study produces an optimal result in terms of sensitivity and accuracy. The results obtained in the present work are comparable and promising. Table 5 shows the comparative analysis of proposed study with similar work carried out by other researchers. Detailed comparative analysis is not possible due to the variability of the size and type of

datasets used in the various related works. Jasionowska et al. used Gabor filter and GLCM-based features to characterize the subtle signs [32]. They achieved a sensitivity of 86 % using the SVM-based classifier. The work suffered from the limitation of not having a large representative and reliable sets of train data. Guo et al. achieved an accuracy of 72.5 % in distinguishing architectural distortion from normal breast parenchyma [5]. Minvathi et al. have achieved 94.78 % sensitivity in distinguishing normal and AD patterns using eight significant features from Haralick model with SVM as a classifier [33]. However, the dataset used in all cases was small, and it included cases of masses showing AD. The classification of pure AD is a more difficult problem due to the subtle and ill-defined appearance. Though, detailed comparative analysis is not possible because of the associated variability of size and type of datasets.

It is quite evident from Table 5 that all the reported values are lower than the ones yielded from our experiments. It is difficult to have a direct and fair comparison as the experimental setup differs. However, our results outperform with accuracy more than 90 %, corresponding to almost all data sets. A very promising value for the sensitivity is achieved. Specificity can be improved in the future by adding representative tumor cases. The program for texture analysis and

Table 5 Comparison of Proposed study with Previous Works

Parameter	Jasionowska et al.[32]	Guo et al.[5]	Minvathi et al.[33]	Proposed method
Database	DDSM	MIAS	MIAS/DDSM	MIAS/DDSM
Dataset	34 AD ROIs, 258 non-AD ROIs.	19 AD ROIs, 21 non-AD ROIs from MIAS.	23 AD ROIs, 97 non-AD ROIs from DDSM. 19 AD ROIs, 152 non-AD ROIs from mini-MIAS.	146 AD ROIs, 75 non-AD ROIs from DDSM. 39 AD ROIs (fixed size). 19 AD ROIs (ground truth) and 108 non-AD ROIs from MIAS.
Accuracy (%)	83.50	72.50	89.69	95.34 (MIAS) 92.94 (DDSM)
Sensitivity (%)	68.00	N.S.	94.38	92.30 (MIAS) 93.33 (DDSM)
Size of ROI (pixels)	Ground truth	128×128	128×128	Fixed size as well as variable size.
Features extracted	GLCM and statistic features	Fractal features	GLCM	GLCM, fractal-based features, Fourier power spectrum features.
Feature selection	Correlation-based feature selection	N.S.	Forward feature selection	Stepwise regression.
Classifier	SVM tuned with kernel parameters	SVM tuned with kernels parameters	SVM and Multi layer perceptrons	SVM tuned with kernel parameters.
Validation	Cross validation (fold not specified).	4 cross validation.	70/30 with leave one out cross validation	3-fold cross validation With testing and training split into 2/3 and 1/3, respectively.
Clinical evaluation	No	No	No	Yes

N.S. not specified in the paper

classification in proposed study is done using MATLAB R2011a on Dell system with Intel Core 2 Duo T7300 (2.0 GHz/4 MB L2 cache), 4GB DDR2 SDRAM@667, and Windows 7 as operating system.

Conclusions and Future Scope

The contributions of the proposed study are threefold. Firstly, most of the feature models showing characterizations with subtle signs are studied collectively to select the optimal ones. Second is the consideration of both the fixed size and variable size ROIs that has given an extensive evaluation for the appropriate selection of size of ROI corresponding to MIAS and DDSM database. Third is the clinical evaluation of current studies on actual patient dataset. The characterization of architectural distortion has been done using texture analysis. Further, by application of stepwise regression to feature set consisting of 284 texture features, the information required for classification of subtle signs is squeezed in 4–5 features. It has been concluded from the result analysis that the selected features corresponding to various datasets have substantially improved the performance of classification. Initially, the samples are divided into AD and non-AD. The performance of the system is validated by comparisons with the other state of the art techniques.

The significant limitations of realized experiments are related to the lack of feature vectors to differentiate between malignant and benign AD. The present work has not concentrated towards the detection of types of architectural distortion viz. peripheral, central, and subareolar. The approach used in the present work can be applied to prior mammograms to enhance the performance of CAD system. The work can be further extended by considering the potential correlation of breast cancer with other types of cancer. It can be done by including all probable types of cancer and analyzing their predictive factors to investigate the correlations, similarities, and differences among them.

Acknowledgments The authors are highly thankful to Dr Ravinder Sidhu, radiologist at ACE Healthways, Ludhiana, India, and Dr Ramandeep Singh, radiologist at Delta Imaging Centre, Ludhiana, India for their support in this work. The authors also thank Dr. Thomas Deserno, Department of Medical Informatics, Aachen University of Technology, Aachen, North Rhine-Westphalia, Germany, for providing the image retrieval in medical application (IRMA) version of DDSM database. The author also thank ACE Healthways, Diagnostic Centre, Ludhiana (India), for providing the actual patient data for carrying out the clinical evaluation of this study. The authors also feel gratitude towards anonymous reviewers for providing substantial and useful review, which led to important improvement in the present manuscript

Conflict of Interest The authors declare that they have no competing interests.

References

1. Breast Cancer in India: Available at <http://global.umich.edu/2013/02/the-growing-problem-of-breast-cancer-in-india>. Accessed 13 January, 2014
2. Population based cancer registry: Available at <http://www.pbcindia.org/>. Accessed 13 January, 2014
3. American College of Radiology: The ACR breast imaging reporting and data system 2003. Available at http://www.acr.org/departments/stand_accred/birads/contents.html. Accessed 25 January, 2014
4. Rao AR, Jain CR: Computerized flow field analysis: oriented texture fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 14:693–709, 1992
5. Guo Q, Shao J, Ruiz V: Investigation of support vector machine for the detection of architectural distortion in mammographic images. *Journal of Physics: Conference Series* 15:88–94, 2005
6. Nandi RJ, Nandi AK, Rangayyan RM, Scutt D: Classification of breast masses in mammograms using genetic programming and feature selection. *Med Biol Eng Comput* 44:683–694, 2006
7. Ayres FJ, Rangayyan RM: Gabor filters and phase portrait for the detection of architectural distortion in mammograms. *Med Biol Eng Comput* 44:883–894, 2006
8. Banik S, Rangayyan RM, Desautels JEL: Computer Aided Detection of Architectural Distortion in Prior Mammograms of Interval Cancer. *J Digit Imaging* 23:611–631, 2010
9. Buciu L, Gacsadi A: Directional features for automatic tumor classification of mammogram images. *Biomed Signal Process Control* 6:370–378, 2001
10. Eltoukhy MM, Faye I, Samir BB: A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multi-resolution representation. *Comput Biol Med* 42:123–128, 2012
11. Harlick RM, Shanmugan K, Dinstein IH: Texture Features for image classification. *IEEE Trans. Syst Man Cybern. Syst.* 3:610–621, 1973
12. Weszka JS, Dyer CR, Rosenfeld A: A Comparative Study of Texture Measures for Terrain Classification. *IEEE Trans. Syst Man Cybern. Syst.* 6:269–285, 1976
13. Mandelbrot BB: *The Fractal Geometry of Nature*. W.H. Freeman and Company, New York, 1986
14. Tourassi GD, Delong DM, Floyd CE: A study on the computerized fractal analysis of architectural distortion in screening mammograms. *Phys. Med Biol* 51:1299–1312, 2006
15. Fukunaga K: *Introduction to Statistical Pattern Recognition*, 2nd edition. Academic, San Diego, 1997
16. Mathworks.com: Available at <http://www.mathworks.com/help/toolbox/Stats/70-71>. Accessed 10 Feb. 2014
17. Cortes C, Vapnik V: Support vector networks. *Mach Learn* 20:273–297, 1995
18. Oliveira JED, Guld M, Araujo A, Ott B, Deserno TM: Towards a standard reference database for computer-aided mammography, *Proceedings of SPIE Medical Imaging* 69151Y, 2008
19. Suckling J, Parker J, Dance J, Astley DR, Hutt S, Boggis I, Ricketts C, Stamatakis I, Cerneaz E, Kok N, Taylor SN, Betal P, Savage D: The Mammographic Image Analysis Society digital mammogram database. *Proceedings of 2nd International Workshop on Digital Mammography*, New York, USA, 1994, pp 375–378
20. Kamra A, Jain VK, Singh S: Extraction of orientation field using Gabor Filter and Gradient based approach for the detection of subtle signs in mammograms. *J. Med. Imaging & Health Infor.* 4:374–381, 2014
21. Li H, Giger ML, Huo Z, Olufunmilayo I, Li O, Barbara L, Weber L, Bonta L: Computerized analysis of mammographic parenchyma

- patterns for assessing breast cancer risk: Effect of ROI size and location. *J. Med. Phys.* 31:549–555, 2004
22. Robert C, Ike III, Singh S, Harrawood B, Tourassi GD: Effect of ROI Size on the Performance of an Information-Theoretic CAD System in Mammography: Multi-size Fusion Analysis. *Proceedings of SPIE, Bellingham, USA, 2008*, p 6915
 23. Bovis K, Singh S: Detection of masses in mammograms using texture features. *Proceedings of the 15th international conference on pattern recognition, Barcelona, Spain, 2000*, pp 267–270
 24. Rangayyan RM, Nguyen TM: Fractal Analysis of contours of breast masses in mammograms. *J Digit Imaging* 20:223–237, 2007
 25. Serrano RC, Conci A, Zamith M, Lima R: About the feasibility of Hurst coefficient in thermal images for early diagnosis of breast diseases. *Proceedings of the 11th Pan-American Congress of Applied Mechanics, Parana, Brazil, 2010*
 26. Theodoridis S, Koutroumbas K: *An Introduction to Pattern Recognition: A MATLAB Approach*, 4th edition. Academic, Burlington, 2010
 27. Sahiner B, Chan HP, Petrick N, Wagner RF, Hadjiiski L: Feature Selection and classifier performance in computer aided diagnosis: the effect of finite sample size. *J. Med. Phys.* 27:1509–1522, 2000
 28. Yang X, Cao A, Song Q, Schaefer G, Su Y: Vicinal support vector classifier using supervised kernel based clustering. *Artif Intell Med* 60:189–196, 2014
 29. Metz CE: Evaluation of digital mammography by ROC analysis. *Proceedings of the 3rd International Workshop on Digital Mammography, Chicago, 1996*, pp 61–68
 30. Witten IH, Frank E: *Data Mining: practical machine learning tools and technique*, 2nd edition. Morgan Kauffman, Burlington, 2005
 31. Cross Validation: Available at http://research.cs.tamu.edu/prism/lectures/iss/iss_113.pdf. Accessed 25 March, 2015
 32. Jasionowska M, Przedaskowski A, Rutczynska A, Wroblewska A: A two step method for detection of architectural distortion in mammograms. *Information Technology in Biomedicine. Advances in Intelligent and Soft Computing.* 69:73–84, 2010
 33. Minavathi, Murali M, Dinesh MS: Model based approach for Detection of Architectural Distortions and Spiculated Masses in Mammograms. *Int. J Comput. Science Engg* 3:3534–3546, 2011