Model based approach for Detection of Architectural Distortions and Spiculated Masses in Mammograms

Minavathi
Information Science and Engineering Department
PES college of Engineering
Mandya, India
minavati@yahoo.com

Murali. S.
Maharaja Institute of Technology
Mandya, India

M. S. Dinesh
PET Research centre
Mandya, India

Abstract— **This paper investigates detection of Architectural Distortions (AD) and spiculated masses in mammograms based on their physical characteristics. We have followed a model based approach which separates the abnormal patterns of AD and spiculated masses from normal breast tissue. The model parameters are retrieved from Gabor filters which characterize the texture features and synthetic patterns were generated using pplanes to retrieve specific patterns of abnormalities in mammographic images. In addition, eight discriminative features are extracted from region of interest (ROI) which describes the patterns representing AD and spiculated masses. Support vector machine (SVM) and Multi-layer Perceptrons (MLP) classifiers are used to classify the discriminative features of AD and spiculated masses from normal breast tissue. This study concentrates on classifying AD and spiculated masses from the ones which actually are normal breast parenchyma. Our proposal is based on the texture pattern that represents salient features of AD and spiculated mass. Once the descriptive features are extracted SVM and MLP classifiers are used. We have used receiver operating characteristic curve (ROC) to evaluate the performance and we have compared our method with several other existing methods. Our method outperformed other existing methods by achieving 90% of sensitivity, 86% specificity in distinguishing AD from normal breast tissue and 93% sensitivity and 88% specificity in classifying spiculated mass from normal breast parenchyma. In first stage of this study we consider ROI's that include AD, spiculated masses and normal breast tissue as input. Our method was tested on 190 ROI's(19 AD , 19 spiculated mass and 152 normal breast tissue) from Mini-MIAS database and 150 ROI's(23 AD , 30 spiculated mass and 97 normal breast tissue) collected from DDSM database. In the second stage we have applied SVM classification model on whole images and the performance is analyzed by plotting Free Response Operating Characteristic (FROC) curves. SVM classifiers achieved 96% sensitivity with 9.6 false positives per image in detection of spiculated mass and 97% sensitivity with 6.6 false positives per image while detecting AD in digital mammograms.**

Keywords- Mammograms, Gabor filters, Architectural distortion, Spiculated masses, Support vector machine, Multi-layer perceptron.

I. INTRODUCTION AND STATE OF ART

Mammography is a most effective imaging modality in early breast cancer detection. The radiographs are searched for signs of abnormality by expert radiologists but mammograms are complex in appearance and signs of early disease are often small or subtle. That's the main reason of many missed diagnoses that can be mainly attributed to human factors [3]. In order to improve the accuracy of interpretation or to prompt with the locations of possible abnormalities, a variety of computer systems have been proposed.

Some of the important signs of breast cancer radiologists normally look for are: spiculated masses, micro calcifications, architectural distortions and bilateral asymmetry. Spiculated masses are characterized by radiating

lines or spicules from a central mass of tissue. Spiculated masses carry a much higher risk of malignancy than calcifications or other types of masses [4]. Architectural distortion is the third most common mammographic finding of breast cancer. Literature says that about 81% of Spiculated mass and 48-60% of AD are malignant and it is estimated that 12-45% of cancers missed in mammographic screening are AD [7]. The detection sensitivity of the current CAD systems for SM and AD is low and there is a pressing need for improvements in their detection. Breast cancer does not always produce a mammographically visible mass, but it frequently disrupts the normal tissues in which it develops. This distortion of architecture may be the only visible evidence of the malignant process. The probability of malignancy increases as a lesion becomes more irregular in shape [5].

In the literature, various methods have been proposed and available for detection of AD and spiculated masses in mammograms. Mudigonda et al. [6] presented a mass detection method that performs segmentation of objects based on isointensity contours and texture flow-field analysis. Their study included 43 masses and 13 normal cases from the Mini-MIAS database with the performance of 81% and average of 2.2 FPs per image. Sampat and Whitman [7] employed filtering in the Radon transform domain to enhance mammograms. They have used radial spiculation filters to detect spiculated lesions. The algorithm was tested on 45 cases exhibiting spiculated masses and on 45 cases with the presence of architectural distortion. A sensitivity of 80% was obtained with 14 false positives per image in the detection of architectural distortion, and 91% with 12 false positives per image in the detection of spiculated masses. Luan Jiang [8] , developed an automated computerized method to detect spiculation levels. A quantitative spiculation index is computed to assess the degree of spiculation. The method achieved an overall classification accuracy of 66.4%, with 54.3% sensitivity and 78.3% specificity. Julia E. E. de Oliveira [15] proposed a method to classify spiculated mass and micro calcification using Haar wavelet transform and SVM. A result of 89.6% of accuracy was achieved by them. Arnau Oliver and Xavier [21] tried to distinguish true spiculated mass from normal breast parenchyma based on local binary patterns and SVM classifiers. They used a set of 1792 suspicious regions of interest extracted from the DDSM database and achieved 90% sensitivity. Leonardo de Oliveira Martins [24] used K-means algorithm and SVM classifiers to detect masses in digital mammograms. Using shape and texture descriptors they classified the masses and obtained the accuracy of 85%. Artificial Neural Networks were used by Mohammed J. Islam and Majid Ahmadi [31] to automatically classify the masses. They used seven features to classify the ROI's and achieved 90.01% sensitivity.

Rangayyan and Ayres [9] applied Gabor filters to characterize oriented texture patterns and detect architectural distortion. The methods were tested with one set of 19 cases of architectural distortion and 41 normal mammograms, and another set of 37 cases of architectural distortion. FROC analysis shows the sensitivity of 0.79 at 8.4 false positives per image. Eltonsy et al. [10] developed a method to detect masses and architectural distortion by locating points surrounded by concentric layers of image activity. The technique was evaluated on 80 images including 13 masses, 38 images with masses and architectural distortion, and 29 images with only architectural distortion. Overall sensitivity of 91.3% with 9.1 false positives per image was obtained.

A Mohd Khuzi et al. and R Besar[11] developed an automated system for assisting the analysis of digital mammograms by extracting the textural features of ROIs by using gray level co-occurrence matrices (GLCM). Results were analyzed plotting ROC curve, where area under ROC rated 0.8 – 0.9 with A_Z= 0.84. M. Arfan Jaffar and Bilal Ahmed [12] have done some experiments for tumor detection in digital mammogram images by extracting eight different multi domain features using SVM and MLP classifiers achieving the accuracy of 85% and 84% respectively. SVM classifiers are used by Guo and Shao [16] along with Hausdorff fractal dimension to detect AD which classified 72.5% of correct answers. S. Baeg and N. Kehtarnavaz [18] worked on mammograms to detect architectural distortions by considering the denseness texture feature. They evaluated their method by plotting ROC and the area under the curve was 0.90.

The detection sensitivity of the current computer systems for spiculated mass and AD is not as good as for microcalcification detection algorithms and there is a pressing need for improvements in their detection. Although significant research effort has gone into developing computerized methods to detect AD and spiculation levels, the accurate and robust detection remains a technical challenge because the spiculated patterns are often subtle and varied in appearance. In addition, because of the lack of "ground truth" spiculation levels in the testing data set, assessing the performance of a method is also difficult. In a way to overcome the shortcomings of existing works we are proposing a method to detect the AD and to minimize the false positives in mammograms.

II. METHODOLOGY

One of the major problems that appear in diagnosis of malignant lesions in mammogram is incorrect classification of lesions. In this paper we propose texture based classification method for Architectural distortions and Spiculated masses in mammograms. Mass is defined as a space occupying lesion seen in at least two different projections. Fat-containing radiolucent and mixed-density circumscribed lesions are benign, whereas isodense to high-density masses may be of benign or malignant origin [20].

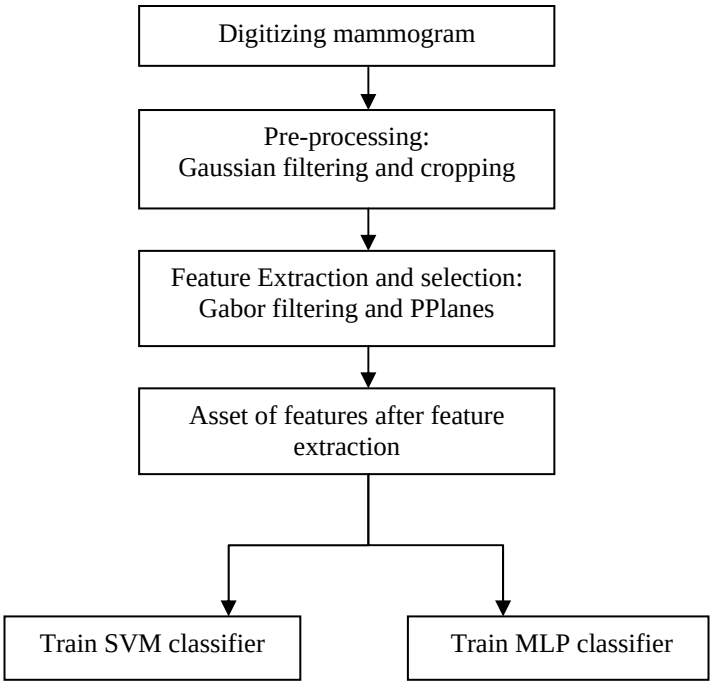

Figure 1. Block diagram showing overall methodology

Benign lesions tend to be isodense or of low density, with very well defined margins and surrounded by a fatty halo, but this is certainly not diagnostic of benignancy. The halo sign is a fine radiolucent line that surrounds circumscribed masses and is highly predictive that the mass is benign. The lesions with spiculated margins are characterized by lines radiating from the margins of a mass shown in Figure 1. A lesion that is ill-defined or spiculated and in which there is no clear history of trauma to suggest hematoma or fat necrosis suggests a malignant process [20].

(a) (b)

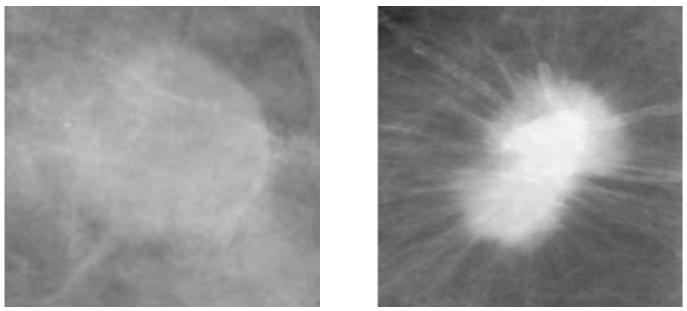

Figure 2. (a) Circumscribed Mass, (b) Spiculated Mass.

Architectural distortions are less prevalent than masses or calcifications, they are the third most common mammographic sign of cancer and are strongly suggestive of malignancy. BI-RADS [29] defined Architectural distortion as “The normal architecture (of the breast) is distorted with no definite mass visible. This includes spiculations radiating from a point and focal retraction or distortion at the edge of the parenchyma. Architectural distortion can also be an associated finding”. Architectural distortion of breast tissue can indicate malignant changes especially when integrated with visible lesions such as mass, asymmetry or calcifications. Architectural distortion can be classified as benign when including scar and soft-tissue damage due to trauma [31]. A mammogram with architectural distortion is shown in Figure 2.

(a) (b)

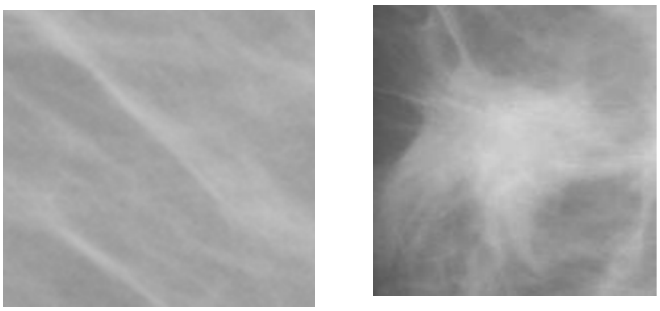

Figure 3. (a) Normal breast structure, (b) Architectural Distorted breast structure

A. Digitizing Mammogram

The overall procedure of this paper is illustrated in the block diagram below. The input mammogram images used in this experiment are from Mini-MIAS database. This database contains left and right breast images total of 161 patients with ages between 50 and 65. All images are digitized at a resolution of 1024 X 1024 pixels and at 8-bit grey scale level. It also includes the locations of any abnormalities that may be present. Our study solely concerns on detection of Architectural distortion and spiculated masses thus, total of 190 mammograms comprising AD, spiculated masses and normal cases were considered. Ground truth information with location and size for these abnormalities are clearly available inside this database.

B. Pre-processing

The original mammograms from Mini-MIAS dataset are (1024 x 1024) pixels, and the whole image is comprised of background with a lot of noise. In the preprocessing step we manually crop the region of image that can approximately cover the region of interest (ROI). Thus, almost all the background information and most of the noise are eliminated. The regions of interest (128 x128) are extracted, where the centers of the abnormality areas are selected to be the centers of ROI. This is done to limit our search for abnormalities without any undue influence from background or unwanted regions of mammograms. To reduce the influence of low-frequency components of mammographic images they are filtered using high pass filter, prior to feature extraction. This is achieved by subtracting the original mammographic image from the Gaussian filtered image. These images are then normalized to improve the quality of the image and to reduce the noise.

C. Feature Extraction and Selection

Feature extraction plays very important role in image processing and also affects the performance of computer aided systems. It is easy to discriminate normal breast tissue from AD or spiculated mass by capturing the significant features like texture feature, gradient based feature and intensity feature. Next important and critical aspect is selection of best set of features. The features considered may include redundant or irrelevant information or the feature taken may not be significant for our classification [33, 34]. It is very important to select features that best suits our classification. Feature selection is to select smaller feature subset from s from set of features f which helps the classifier to perform well.

The distribution of the mammary gland is approximated to linear structures. In a normal breast, the direction of the distribution tends toward the nipple and in an abnormal breast it tends toward suspect areas as shown in figure 1 and 2. As a first set of characteristic features we consider gray level primitives that constitute the image texture. Based on the linear structure of mammary gland, we evaluate local structure of mammary ducts. In the present work we have used Gabor filters as line detectors. In order to extract the texture orientation at each pixel of a mammogram, we filter the mammogram with a bank of Gabor filters of different orientations [9]. Gabor filters provide good detection accuracy for linear patterns. The Gabor filter kernel oriented at the angle $\theta = -\pi/2$ is given by:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right] \quad (1)$$

Kernels at other angels can be obtained by rotating this kernel. We have used 180 kernels with angles spaced evenly over the range $[-\pi/2, \pi/2]$. In order to extract orientation at each pixel of magnitude image obtained after Gabor filtering, gradient based orientation extraction is used. Here gradient vectors are calculated by taking partial derivatives of image intensity at each pixel in Cartesian coordinates.

1) Architectural distortion

Kopans [5] has pointed out that: “Most breast cancers are radio graphically very dense for their size and high in x-ray attenuation they are generally not intermingled with fat and their x-ray attenuation increases toward their canters". Process for detecting AD is further performed by searching for node like structures in the image. Phase Planes provide an analytical tool to study systems of first-order differential equations.

(a) (b)

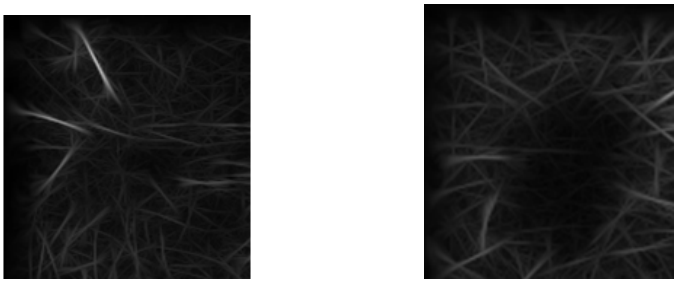

Figure 4. (a) Magnitude image of AD , (b) Magnitude image of Spiculated mass after Gabor filtering.

The phase portrait of a system of differential equations is the graphical representation of the possible trajectories in the phase plane. Node maps are drawn using phase planes (PPlanes) and it is compared with the gradient orientation image of previous section. We need some measure of distance between two orientation fields as we are dealing with orientation fields. Flow field is analyzed using Distance measure (nonlinear least squares) [2, 35] and the degree of distortion is calculated. The measure that we use is the area of the triangle formed by the oriented segments, which is

$$A_{i,j} = \frac{1}{2} R_1 R_2 \left| \sin(\theta_{1_{i,j}} - \theta_{2_{i,j}}) \right| \quad (2)$$

We have considered two discrete orientation fields O1 and O2 where O1 is the orientation extracted from node map drawn from PPlanes and O2 is the gradient orientation of mammographic image. Length of orientation line segment is represented by $R_{k_{i,j}}$ in orientation field O_k at location (i, j), θk be the angle subtended by this line segment. Area $A_{i,j}$ is a measure of the difference or disparity between two oriented segments. When two segments have the same orientation the area becomes zero. We find the sum of these differences over the entire field to obtain

$$S = \sum_{i,j \in W} A_{i,j} \quad (3)$$

As S becomes smaller two fields will be closer. The presence of strong node point indicates the sites of architectural distortion. The first set of primitives features that constitute image texture extracted above are not sufficient to classify them as AD or normal. Set of second characteristic features that are concerned with spatial organization of gray level primitives are considered. We have extracted the following features from gradient orientation image: mean, variance, standard deviation, skewness, entropy, energy, contrast and homogeneity to discriminate normal images from AD.

(a) (b)

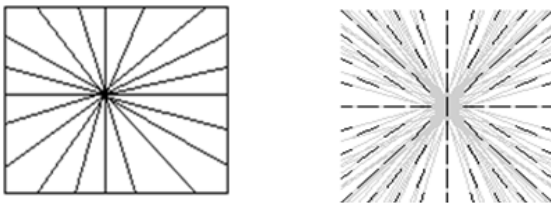

Figure 5. (a) Node map drawn using PPlanes, (b) Orientations indicating nodal patterns

2) Spiculated Mass

Spiculated masses have stellate appearance as shown in figure 6 below. The size of the lesion ranges from few millimetres to centimetres. Along with the texture feature that is retrieved from Gabor filters and the orientation features retrieved from gradient orientation method we extract second set of features that are relevant to stellate appearance of mass.

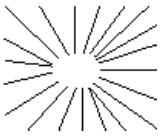

Figure 6. Shows stellate appearance of spiculated mass

Texture features are useful in discriminating spiculated mass from normal images. Such texture descriptive discriminating features are extracted from Gray level co-occurrence matrices (GLCM). Features that are extracted from GLCM for detection of spiculated masses are energy, homogeneity, entropy, correlation and contrast.

D. Classification

Out of various typical non-linear classifiers, two popular representative neural architectures are considered. Multi-layer Perceptrons and Support vector machine are used to classify the ROI's including AD, Spiculated mass and Normal breast tissue. The main task of the classifier is to categorize the ROI by identifying the probability for each of the possible categories. The process of classification has two phases: training phase and testing phase. In training phase data set which is labelled as Normal, AD or Spiculated mass are given to classifier and the classifier is trained. Whereas in testing phase, unknown dataset are given to classifier for actual classification.

1) Multi-Layer Perceptrons (MLP)

Multilayer-Perceptrons are popular supervised training neural network models that are used to train the network. MLP contains three layers namely: input, output and hidden layers. Input nodes, output nodes and hidden nodes are connected via variable weights using feed forward connections [38]. But when output error occurs, it is difficult to say how much error comes from different nodes and there will be problem in adjusting weights according to their contribution. Back-propagation algorithm is used to solve this. Feature reduction was not required for neural network classifiers because the neurons of input layer (trained) will be examined to discard the features. Final output will be compared with target output and total mean square error is calculated using all training patterns of calculated and target outputs.

2) Support vector machine (SVM)

SVM is a learning tool based on modern statistical learning method that classifies binary classes. SVM finds and uses class boundary hyper plane by maximizing the margin in training data. The training data samples along the hyper planes near the class boundary are called support vectors. Using support vectors SVM finds adequate hyper plane to separate the groups. After separation cases belonging to one category remains in one side of the plane and other cases on the other side of the plane [36].

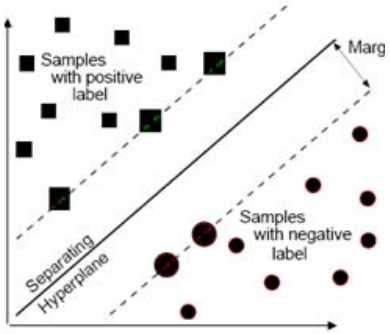

Figure 7. Separation of two classes by an optimal hyper plane.

The reason why SVM is selected for classification in our work is [37]:

a) SVM has good capacity of generalization.
b) SVM is highly robust and work well with images.
c) The theory of SVM is well defined and has a very good base of mathematics and statistics.
d) Over training problem is less compared to other neural network classifiers.

Let us denote the training set containing n samples as : $\{(x_i, y_i), \quad i = 1,2 \ldots\ldots, m)\}$

Each instance in the training set contains one "target value" (class labels) and several "attributes" (features). Vector x $\in R^n$ denote pattern to be classified to either of two classes labeled $y_i \in (-1)$. It maps x to higher dimensional space H using nonlinear operator $\Phi(\bullet) : R^n \rightarrow H$. Hyper plane which separated the classes of input patterns is constructed as:

$$f(x) = w^T \Phi(x) + b, w \in H, b \in R \qquad (4)$$

The aim of SVM is to minimize $\|w\|$ and to separate the data with minimum number of errors. Mathematically it can be stated as: $min\left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \varepsilon_i\right]$ (5)

Where $\varepsilon_i \geq 0$ and satisfies the constraint:

$$y_i(w^T\Phi(x_i) + b) \geq 1 - \epsilon_i \qquad (6)$$

and i = 1,2… m

Parameter C controls the penalty for misclassifying the training samples. We can transform this optimization problem to a dual form using the Lagrange multiplier technique. This transformation also shows that, necessary condition for minimizing (5) is to form vector w as follows:

$$w = \sum_{i=1}^{m} \alpha_i y_i \Phi(x_i) \qquad (7)$$

Where $\alpha_i \geq 0$, i = 1,2 .... m, is a vector of Lagrange multipliers and they are solved using dual form given by:

Maximize; $w(\alpha_1, \alpha_2 \ldots\ldots \alpha_m) = \sum_{i=1}^{m} \alpha - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j k(x_i, x_j)$ (8)

Subjected to : $0 \leq \alpha_i \leq C$, i = 1,2 ….. m

$$\sum_{i=1}^{m} \alpha_i y_i = 0$$

Kernel function k(x, y) in SVM plays very important role in mapping input vector to high dimensional feature space. By using different kernel functions like Gaussian Radial Basis Function (RBF), Polynomial, Linear, quadratic, etc., SVM implements a variety of learning machines.

These kernel functions are defined as follows:

Gaussian RBF kernel: $k(x, y) = exp\left[-\frac{\|x-y\|^2}{\sigma}\right]$ (9)

Polynomial kernel: $k(x, y) = ((x \cdot y) + coeff)^p$ (10)

Linear kernel : $k(x, y) = (x \cdot y)$ (11)

Where $\sigma \in R$, width of Gaussian RBF function and $P \in N$, degree of polynomial function.

III. EXPERIMENTS AND RESULTS

Our work is being carried out using Mini-MIAS and DDSM database. The database provides mammograms with ground truth information. Implementation is done using Mat lab R2009a. Images were stored in uncompressed gray-scale PNG files. We have worked on total of 340 images where 190 images (19 AD, 19 spiculated mass and 152 normal images) were from Mini-MIAS database and 150 images (23 AD 30 spiculated mass and 97 normal images) from DDSM database. First set of image feature vectors were extracted using

Gabor filters as explained in section 2.3 and next set of features were retrieved form gradient orientation images for AD and from GLCM for Spiculated masses as explained in 2.3.1 and 2.3.2. Classification is done using SVM and MLP classifiers using SVM toolbox and neural network toolbox. SVM are widely used in Medical image processing which provides a numerical value related to the membership of each class. ROC curve can be generated by verifying the threshold of this membership. SVM classifiers were evaluated using linear kernel, Polynomial kernel and Gaussian RBF kernels.

Dataset was divided as Training set 70% and Testing set 30%. We have evaluated the method of classification for two breast abnormalities AD and Spiculated mass, without differentiating them as benign or malignant. We have randomly selected images from Mini-MIAS database. As the images having abnormalities AD and spiculated mass in Mini-MIAS dataset (i.e. 19 AD and 19 spiculated mass) were not sufficient for our classifiers, we carried out our test even on DDSM dataset. We have used Forward feature selection method to find the best features. For cross validation we used leave-one-out scheme. In this method decision is done using all but one of samples in the dataset. The left out sample is then used to test the decision rule performance. Performance analysis is done by plotting Receiver Operating Curve (ROC). ROC graphically represents the true positive rate as a function of false positives rate. The performance of SVM for 3 kernel functions is given in table 1 and table 2.the performance of the classifier is assessed in terms of sensitivity and specificity. Where sensitivity is the proportion of actual positives which are correctly identified and specificity is the proportion of negatives which are correctly classified. The sensitivity achieved by SVM classifier for classifying AD and Normal images is 94.7% and for spiculated mass and Normal images is 93.05%. Performance of our method consisting of Gabor filter, eight discriminating features for SVM and MLP classifiers is seen to be better than other existing methods which are shown in table 4. Table 3 gives the results of MLP classifier for AD and spiculated mass. It best performed for classifying spiculated masses giving 91.17% sensitivity. ROC curve depicting the performance of SVM classifier for AD with sensitivity 94.7%, area under curve (AUC) is 0.90 and ROC curve of SVM classifier for spiculated mass with sensitivity 93.05%, AUC is 0.83 is shown in figure 8. The performance of MLP classifier is plotted using ROC curve in figure 9. ROC is plotted for classifying AD and normal images and the sensitivity is 89.06 % for spiculated mass and normal images sensitivity is 91.17%.

TABLE I. RESULTS OF SVM CLASSIFIER FOR AD AND NORMAL IMAGES

Kernel Function	Sensitivity (%)	Specificity (%)
Gaussian RBF	94.78	89.67
Polynomial	85.14	72.39
Linear	84.82	80.15

TABLE II. RESULTS OF SVM CLASSIFIER FOR SPICULATED MASS AND NORMAL IMAGES

Kernel Function	Sensitivity (%)	Specificity (%)
Gaussian RBF	93.05	87.92
Polynomial	82.51	65.65
Linear	86.67	79.41

TABLE III. RESULTS OF MLP CLASSIFIER

Case	Sensitivity (%)	Specificity (%)
AD	89.06	76.8
Spiculated Mass	91.17	84.6

TABLE IV. COMPARISION OF OUR METHOD WITH OTHER EXISTING METHODS

Technique	Problem addressed	Sensitivity (%)
Radial spiculated filter [7]	Mass	91
Artificial neural network [38]	Mass	90.91
Multi resolution and SVM [13]	Mass	80
Haar wavelet and SVM [15]	Mass	89.6
Local binary patterns and SVM [21]	Mass	90
K-means and SVM [24]	Mass	86.1
Decision tree and SVM [25]	Mass	85
Hausdorff fractal dimension and SVM classifier [16]	AD	72.5
Denseness feature and neural network [18]	AD	90
Radial spiculated filter [7]	AD	80
Gabor filter and SVM (Our method)	Mass	93.05
Gabor filter and MLP(Our method)	Mass	91.17
Gabor filter and SVM (Our method)	AD	94.78
Gabor filter and MLP(Our method)	AD	89.06

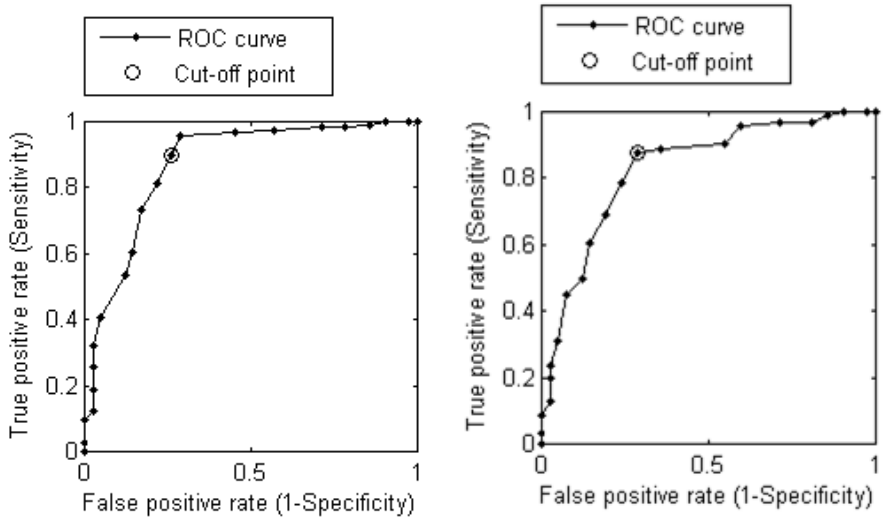

Figure 8. ROC curve depicting the performance of SVM classifier for AD with sensitivity 94.78%, AUC=0.90. ROC curve of SVM classifier for spiculated mass with sensitivity 93.05, AUC=0.83.

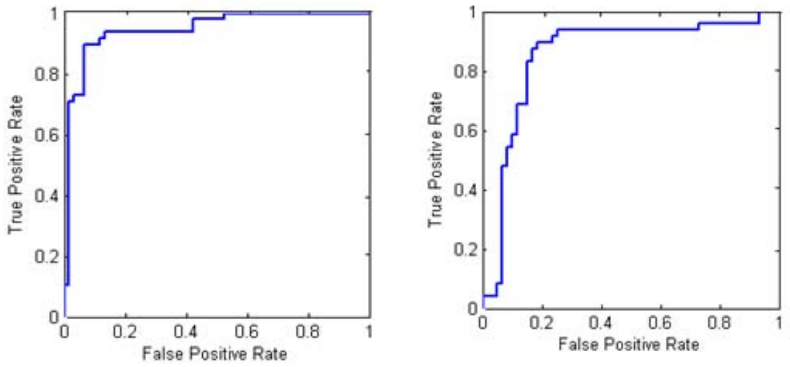

Figure 9. ROC curve depicting the performance of MLP classifier for AD with sensitivity 89.06 % and ROC curve of MLP classifier for spiculated mass with sensitivity 91.17% respectively.

A. Detection of AD and Spiculated Mass

Classifiers which were used to classify normal breast tissue from AD and spiculated masses encouraged us to apply the method of classifying ROI's on whole mammogram. The model built from training set was then applied on large dataset of mammograms to determine the overall performance of detecting the abnormalities as follows.

Step1. Mammograms having abnormalities AD and spiculated masses are first preprocessed for removal of noise as explained in our previous paper [1]. Preprocessing is done using connected component method and anisotropic diffusion. Connected component algorithm was useful to effectively remove artifacts in the mammographic images. The pectoral muscle represents a predominant density region in most medio-lateral oblique (MLO) views of mammograms, and can affect the results of image processing methods. Anisotropic diffusion method is used to smooth the homogenous areas of the image while enhancing the edges. Diffusion model not only provides different degrees of smoothing for intra-regions but also actively provides different degrees of sharpening for edges in inter-regions. After identifying the pectoral region it is removed using Region growing method [1].

Step 2. A window of 50 x 50 is slide on the whole image. The classification model using SVM classifier which is discussed in section 2 is applied on each window considering each 50 x 50 image. Classified true positives are retained in the image indicating the suspected area of abnormality omitting other windows. Figure 10 and 11 shows example of applying the technique to a single mammogram that contains AD and spiculated mass respectively, which is clearly highlighted in the probability image.

Performance of this is measured using Free Response Operating Characteristic (FROC) curves a variant of a receiver-operating characteristic (ROC). FROC is a plot of operating points which shows the tradeoff between the true positive rate and the number of false positives per image, varying the threshold applied to the probability image. Threshold considered allows us to detect true positive for each window if it at least partially overlapped the abnormality, else it is recorded as false positive. The goal of almost all screening algorithms is to maximize the percentage of masses found (sensitivity), while minimizing the number of false positives per image. Figure 12 shows the FROC plots for detecting spiculated mass and AD using SVM classifiers. The performance measure achieved by SVM classifiers for detecting spiculated mass is as follows: Sensitivities of 92%, 94% and 96% with 4.4,7.2 and 9.6 false positives per image respectively. SVM classifiers for detection of AD resulted in achieving the sensitivities of 89% , 95% and 97% with 3.8, 5 and 6.6 false positives per image respectively.

IV. CONCLUSIONS

Architectural distortion is a third main abnormality in mammograms and is often overlooked by radiologists. Both AD and spiculated mass cause mammary distortion due to intrusion of cancer. They play very important role in discriminating benign and malignancy. In this paper we proposed method to classify AD and speculated mass from normal breast using Gabor filters, SVM and MLP classifier. Gabor filter used in this work helped in extracting line structures which in turn supported us to retrieve node and star patterns in image. Other eight selected features provided discriminating characteristics sufficient enough to classify AD and spiculated mass from normal breast ROI's. MLP and SVM classifiers were used and SVM provided nearly 94.7 % sensitivity in classifying AD from normal images and 93% sensitivity in classifying spiculated mass from normal breast tissue. Our method showed good performance compared to existing methods as shown in table 4. We have also continued applying SVM classification model on a whole mammographic image and the performance is analyzed by plotting FROC curves. We have achieved a sensitivity of 96% with 9.6 false positives per image in detection of spiculated mass and 97% sensitivity with 6.6 false positives per image while detecting AD in digital mammograms. Future research will concentrate on designing image processing algorithms to extract features from different modalities and combining these features to improve the performance of detection and classification of AD and spiculated mass.

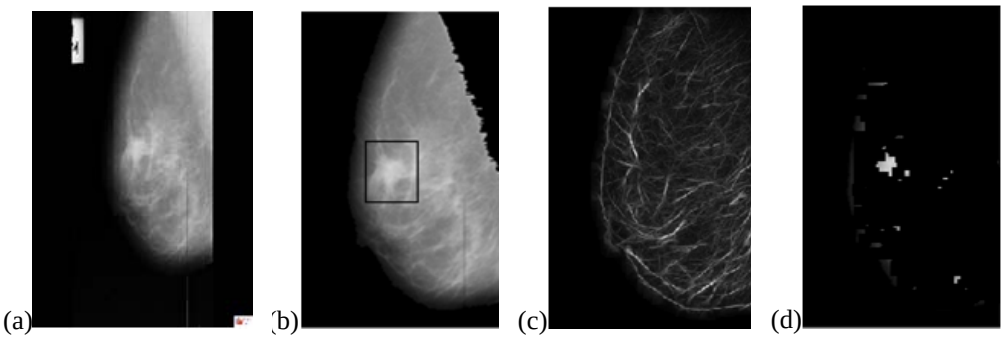

Figure 10. (a) Original mammogram , (b) Pre-processed image showing the position of AD, (c) line strength image after applying Gabor filter, (d) Probability image

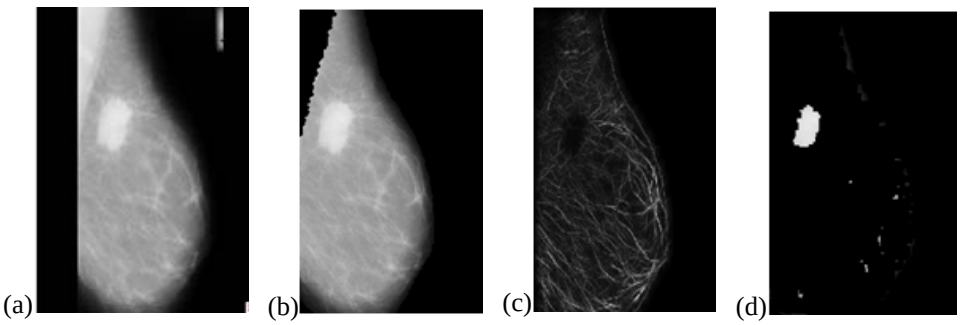

Figure 11. (a) Original mammogram , (b) Pre-processed image showing the position of Spiculated mass, (c) line strength image after applying Gabor filter, (d) Probability image.

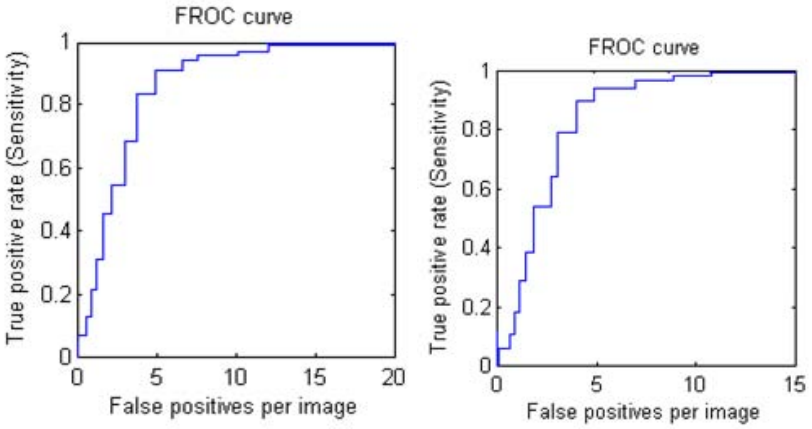

Figure 1. FROC curves for detection of Spiculated mass and FROC curves for detection of AD using SVM classifiers respectively.

REFERENCES

[1] Minavathi, Murali. S , M. S. Dinesh, " Segmentation scheme for mammograms ", Proceedings of ICETE, Nitte, 210-214, 2011.

[2] Minavathi, Murali. S , M. S. Dinesh, "Detection of Architectural Distortions with Spiculations in Mammograms by analyzing the structure of mammary glands " , Accepted in Fifth Indian International Conference on Artificial Intelligence(IICAI) , Tumkur, tobe held in Dec-2011.

[3] Ranadhir Ghosh, Moumita Ghosh, John Yearwood, (2004) ,"A Modular Framework for Multi category feature selection in Digital mammography", ESANN'2004 proceedings, ISBN 2-930307-04-8, pp. 175-180 [2].

[4] Mehul P. Sampat,Gary J. Whitman, Alan C. Bovik, and Mia K. Markey, " Comparison of Algorithms to Enhance Spicules of Spiculated Masses on Mammography", Journal of Digital Imaging, Vol 0, No 0 (Month), 2007: pp 1-8.

[5] D. Kopans," Breast Imaging", New York: Lippincott-Raven Publishers, 1998.

[6] N. R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels, "Detection of breast masses in Mammograms by density slicing and texture flow-field analysis", *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1215–1227, December 2001.

[7] Sampat MP,Whitman GJ, MarkeyMK, BovikAC (2005) , "Evidence based detection of spiculated masses and architectural distortion". Int : Fitzpatrick JM, Reinhardt JM (eds) Proceedings of SPIE medical imaging 2005: image processing, vol. 5747, San Diego, pp 26–37.

[8] Luan Jiang, Enmin Song, Xiangyang Xu, Guangzhi , Bin Zheng , "Automated Detection of Breast Mass Spiculation Levels and Evaluation of Scheme Performance1 ", Acadamic Radiology , Vol 15 , No 12, Dec 2008, 15:1534–1544.

[9] Rangayyan RM, Ayres FJ (2008) , "Detection of architectural distortion in prior screening mammograms using Gabor filters, phase portraits, fractal dimension, and texture analysis", Int JCARS (2008) 2:347–361 .

[10] Eltonsy N, Tourassi G, Elmaghraby A (2006), " Investigating performance of a morphology-based CAD scheme in detecting architectural distortion in screeningmammograms ", Int Lemke HU,Inamura K, Doi K, Vannier MW, Farman AG (eds) Proceedings of the 20th international congress and exhibition on computer assisted radiology and surgery (CARS 2006), Osaka, Japan, June 2006. Springer, Heidelberg, pp 336–338.

[11] A Mohd. Khuzi, R Besar, WMD Wan Zaki, NN Ahmad, "Identification of masses in digital mammogram using gray level co-occurrence matrices", Biomed Imaging Interv J 2009; 5(3): e17.

[12] M. Arfan Jaffar, Bilal Ahmed, Ayyaz Hussain, "Multi domain Features based Classification of Mammogram Images using SVM and MLP", Fourth International Conference on Innovative Computing, 978-0-7695-3873-0/09 , 2009 IEEE.

[13] Campanini R, Dongiovanni D, Iampieri E, Lanconelli N, Masotti M, Palermo G, Riccardi A, Roffilli (2004). "A novel featureless approach to mass detection in digital mammograms based on support vector machines", Phys. Med. Biol., 49: 961-975.

[14] Ayres FJ, Rangayyan RM (2005), " Characterization of architectural distortion in mammograms " , IEEE EngMed Biol Mag 24(1):59–67.

[15] Julia E. E. de Oliveira , Thomas M. Deserno and Arnaldo de A. Araujo, " Breast Lesions Classification applied to a reference Database ", IEEE and E-MEDISYS 2008 2nd International Conference: E- Medical Systems.

[16] Guo Q, Shao J, Ruiz V (2005), " Investigation of support vector machine for the detection of architectural distortion in mammographic images, " J. Phys. Conf. Ser., 15: 88–94.

[17] R.M. Nishikawa , " Current status and future directions of computer-aided diagnosis in mammography" , Computerized Medical Imaging and Graphics, 31(4-5):224– 235, 2007.

[18] S. Baeg and N. Kehtarnavaz, " Classification of Breast Mass Abnormalities using Denseness and Architectural Distortion ", ELCVIA ISSN: 1577-5097, 2002.

[19] R. Gupta and P. E. Undrill, "The use of texture analysis to delineate suspicious masses in mammography," Phys. Med. Biol., vol. 40, pp. 835–855, 1995.

[20] E.S. de Paredes: Atlas of Mammography, 3rd Edition, Lippincott Williams & Wilkins, Philadelphia, USA, 2007.

[21] Arnau Oliver, Xavier Llado, Jordi Freixenet, and Joan Mart, " False Positive Reduction in Mammographic Mass Detection Using Local Binary Patterns ", Springer-Verlag Berlin Heidelberg 2007, pp. 286–293, 2007.

[22] N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms," Phys. Med. Biol., vol. 43, no. 2, pp. 365–378, Feb. 1998.

[23] Dave Tahmoush , " Image Similarity to Improve the Classification of Breast Cancer Images ", Algorithms , ISSN 1999-4893, 2009.

[24] Leonardo de Oliveira Martins, Geraldo Braz Junior, Arist ofanes Correa Silva, Anselmo Cardoso de Paiva, and Marcelo Gattass, " Detection of Masses in Digital Mammograms using K-means and Support Vector Machine ",ELCVIA ISSN:1577-5097, 8(2):39-50, July 2009.

[25] Medhat Mohamed Ahmed Abdelaal, Muhamed Wael Farouq, Hala Abou Sena and Abdel-Badeeh Mohamed Salem, "Applied Classification Support Vector Machine for providing Second Opinion of Breast Cancer Diagnosis ", The Online Journal on Mathematics and Statistics (OJMS) Vol. (1) – No. (1), Reference Number: W10-0010 1, 2010.

[26] Mehul P. Sampat, Alan C. Bovik, "Toroidal gaussian filters for detection and extraction of properties of spiculated masses", ICASSP 2006, 142440469X/ 06/, 2006 IEEE , pp593-596.
[27] Jinshan Tang, Rangaraj M. Rangayyan , Jun Xu, Issam El Naqa, and Yongyi Yang , "Computer- Aided Detection and Diagnosisof Breast Cancer With Mammography: Recent Advances" , IEEE transactions on Information technology in biomedicine, vol. 13, no. 2, march 2009.
[28] Sheng-Fang Huang, Ruey-Feng Chang, "Characterization of Spiculation on Ultrasound Lesions ", IEEE transactions on Medical Imaging, vol. 23, no. 1, January 2004, pp111-121.
[29] American College of Radiology (ACR) (1998) , " Illustrated breast imaging reporting and data system", (BI-RADS), 3rd edn , American College of Radiology, Reston.
[30] H. Burrell et al., False-negative breast screening assessment. What lessons can we learn? Clinical Radiology 56 (5), 385 (2001).
[31] Tomasz Arodz, Marcin Kurdziel, " Pattern recognition techniques for automatic detection of suspicious-looking anomalies in mammograms ", Computer Methods and Programs in Biomedicine (2005) Elsevier Ireland, 135—149.
[32] Jelena Bozek, Kresimir Delac, "Computer-Aided Detection and Diagnosis of Breast Abnormalities in Digital Mammography", 50th International Symposium ELMAR-2008, 45-52.
[33] Gonzalez R.C., and Woods R.E., Digital Image Processing (Second Ed), Prentice Hall, 2002.
[34] O.Duda, P.E Hart, and D,G,stork, Pattern Classification,2nd ed. New York:Wiley – Inter science, 2000.
[35] A. Ravishankar Rao and Ramesh C, " Computerized Flow Field Analysis: Oriented Texture Fields" , 0162-8828 , 1992 IEEE.
[36] Vapnik V 1998 Statistical learning theory (NY: Wiley).
[37] Steve R. Gunn, "Support Vector Machines for Classification and Regression ",Technical Report, University of Southampton, 1998.
[38] Mohammed J. Islam, Majid Ahmadi, "An efficient automatic mass classification method in digitized mammograms using Artificial neural network " , IJAIA, Vol.1, No.3, July 2010.