Hybrid Texture based Classification of Breast Mammograms using Adaboost Classifier

M. Arfan Jaffar

College of Computer and Information Sciences,
Al Imam Mohammad Ibn Saud Islamic University (IMSIU),
Riyadh, Saudi Arabia

Abstract—**Breast cancer is one of the most dangerous, leading and widespread cancers in the world especially in women. For breast analysis, digital mammography is the most suitable tool used to take mammograms for detection of cancer. It has been proved in the literature that if it can be detected at early and initial stages, then there are many chances to cure timely and efficiently. Therefore, initial screening of mammograms is the most important to detect cancer at initial stages. A radiologist is very expensive in the whole world wide and for a common person, it is very difficult to take opinion from more than one radiologist because it is a very sensitive disease. Thus, another solution is required that can be used as a second opinion to help the low cost solution to the patients. In this paper, a solution has been proposed to solve such type of problem to take mammograms and then detect cancer automatically in those images without any help of radiologist or medical specialist. So this solution can be adopted especially at the initial level. Proposed method first segment the portion of the image that contains these cancerous parts. After that, enhancement has been performed so that cancer can be clearly visible and identifiable. Texture features have been extracted to classify mammograms. An ensemble classifier AdaBoost has been used to classify those features by using the concept of intelligent experts. The standard dataset has been used for validation of the proposed method by using well-known quantitative measures. Proposed method has been compared with the existing method. Results show that proposed method has achieved 96.74% accuracy as well as 98.34% sensitivity.**

Keywords—Features; Segmentation; Breast mammograms; Classification; Texture

I. INTRODUCTION

Cancer is the most dangerous and leading cause of death in the whole world wide. There are different types of cancers in the different organs of a human. For women, breast is the most important organ. At the baby birth, mother women used his breast to feed her milk to her child. Therefore, breast is the most important organ especially for women. There is a special care required so that it can escape from any type of disease or cancer. Due to milk transfer to child's, there are chances that may be cancer also shifted to child's if it is uncured or due to unaware of such type of diseases [1]. Breast cancer is the most common cancer especially in the women. Thus there is special attention required to solve this problem. Mammography is a process that can be used to detect cancer in the breast. Radiologists are the most expensive in the whole world wide. It is very difficult for a common person to bear too many expenses. Second this cancer is also diagnosed very carefully.

Most of the time, it is recommended to take the second opinion from another radiologist. Due to lack of funds or expenses, it is very difficult to take the second opinion. Now a day, in this digital world, it is possible to introduce a computer based solution to diagnose such type of cancers [2, 3, 4]. In the literature, many different Computer Aided Diagnosis (CAD) systems are available to help the radiologist to take the second opinion. Most of the existing system has some problems due to poor imaging quality. Some systems did not perform well in the case of noises or due to low radiation may be image has low quality or poor quality due to low contrast. There is CAD system available that guarantees the solution. Still, there is room to improve the performance of these CAD systems [5,6]. Therefore, I have tried to propose a solution to detect cancer in the breast mammograms.

In this paper, a new CAD system has been proposed by using different three types of steps. First breast part of the mammograms has been extracted by using a bilateral filter with logarithm transformation. This bilateral filter smooths the gray levels by preserving the edges. Log transformation has an advantage that it increases the dynamic range especially for those areas which are dark in the mammograms. Then entropy has been calculated so that thresholding can be applied to make it binary. Then seed point has been selected from the white area so that adaptive contour method can start. After extracting breast part, enhancement has been performed to improve the performance. Then features extraction has been performed to classify busing ensemble classifier.

The main contribution of the proposed methods is following:

- Proposed method works well for low contrast images as well due to bilateral filter, log transformation and enhancement process.

- Adaptive contour method has been used by using the concept of entropy with active contour.

- Enhancement has been performed by using Partitioned Iterated Function System.

- The classification has been performed by ensemble classifier AdaBoost.

II. PROPOSED METHOD

Proposed method consists of different phases to complete the whole process. Figure 1 shows sample image from the

dataset and it clearly shows that this mammogram image has many different parts inside it. There is some portion that is background, some portion shows muscles that are not part of the area where we have to find out cancer. Therefore, it is important to remove all these unwanted parts and segment the required part for further analysis. In the first phase, segmentation has been performed to extract the region of interest. In the second phase, there is also required to enhance the quality of the image so that it can be shown clearly visible and easily identifiable. Thus enhancement has been performed to improve the quality of mammograms. After that, features are required to classify those regions. Figure 1 shows that mammograms clearly show texture on the image. So Texture features have been extracted and later used for classification. We know that in our daily life different experts can give their opinion and finally conclusion has been designed by combining the opinions of all those experts. A similar concept has been used in this paper to classify mammograms. Ensemble classifier AdaBoost has been used that can combine the performance of different classifiers and finally decide the output by using the texture features. Details of all these phases has been given below in detail
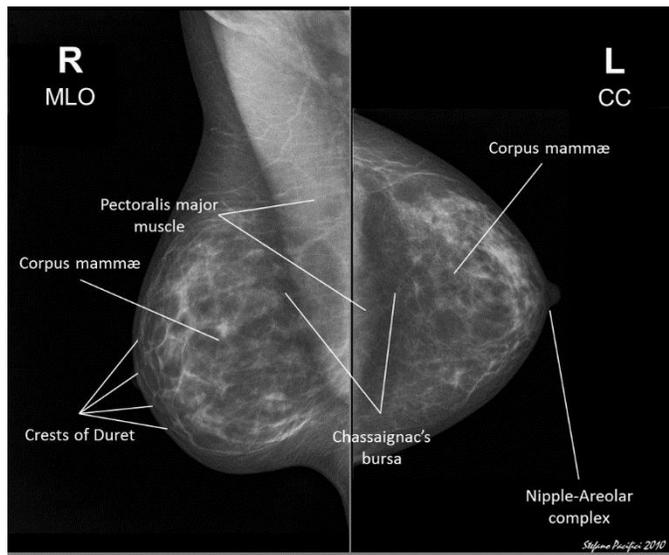
Fig. 1.   Left and Right Mammogram Image with labelling [15]

TABLE I.   List of Abbreviation used in this Paper

CAD	Computer Aided Diagnosis
Log Transformation	Logarithmic Transformation
SVM	Support Vector Machine
KNN	K Nearest Neighbour
ANN	Artificial Neural Network

A. Preprocessing

In this phase, segmentation has been performed to extract the portion of the region of interest that can be used later for

features extraction and classification. For segmentation, adaptive active contour method has been used to segment automatically. In the literature, it many researchers has used active contour but the major problem with active contour is the seed point where it needs to start. Therefore, in this paper, I have modified the existing active contour that works based upon snake by using the concept of entropy. Entropy can be used to find the area where this active contour needs to start. Further, to improve the performance, the bilateral filter has been applied at the start so that edges can be preserved. Therefore, first images have been filtered by using bilateral filter [7] so that this filter performs smoothing on the images while it also preserves the edges.

$$y(m,n) = \sum_k \sum_l h[m,n;k,l]x[k,l] \qquad (1)$$

After preserving the edges and making the image smoother, logarithm transformation [8] has been applied to the image as shown in Figure 2. The basic advantage of log transformation is to increase the dynamic range especially in those areas which are dark in the mammograms as shown in Figure 3.

$$I_o = c \ln[1 + (e^\sigma - 1)I_{in}(i,j)] \qquad (2)$$

Where Io is output image, c is constant, Iin is input image and σ is the scaling factor that controls the input range to the logarithmic function. So this log transformation can also improve the low contrast areas and regions available inside the images. After this step, active contour has to apply for the segmentation of breast part [9]. For active contour, a seed point is required and this seed point can find out by using the concept of entropy. So entropy has been calculated and applied to the image as a threshold. After making threshold the image, a seed point can be selected that is the white area which shows the breast part of the image. After applying a threshold, the boundary of the breast is not accurate due to overlapping intensities of inside breast part and outer side. But at least a point can be used to start the active contour.

Active contour works on the base of the snake. The concept of active contour has been taken from [9]. First energy is required to calculate that can be calculated by using an energy function shown in equation (3).

$$E_{snake} = \int_0^1 E_{snake}(v(s))ds$$

$$E_{snake} = \int_0^1 E_{snake}(v(s)) + E_{ext}(v(s))ds \qquad (3)$$

$$E_{snake} = \int_0^1 E_{int}(v(s)) + E_{img}(v(s))$$
$$+ E_{con}(v(s))ds$$

Before Bilateral Filter   After applying Bilateral Filter

(a)

Before Bilateral Filter   After applying Bilateral Filter

(b)
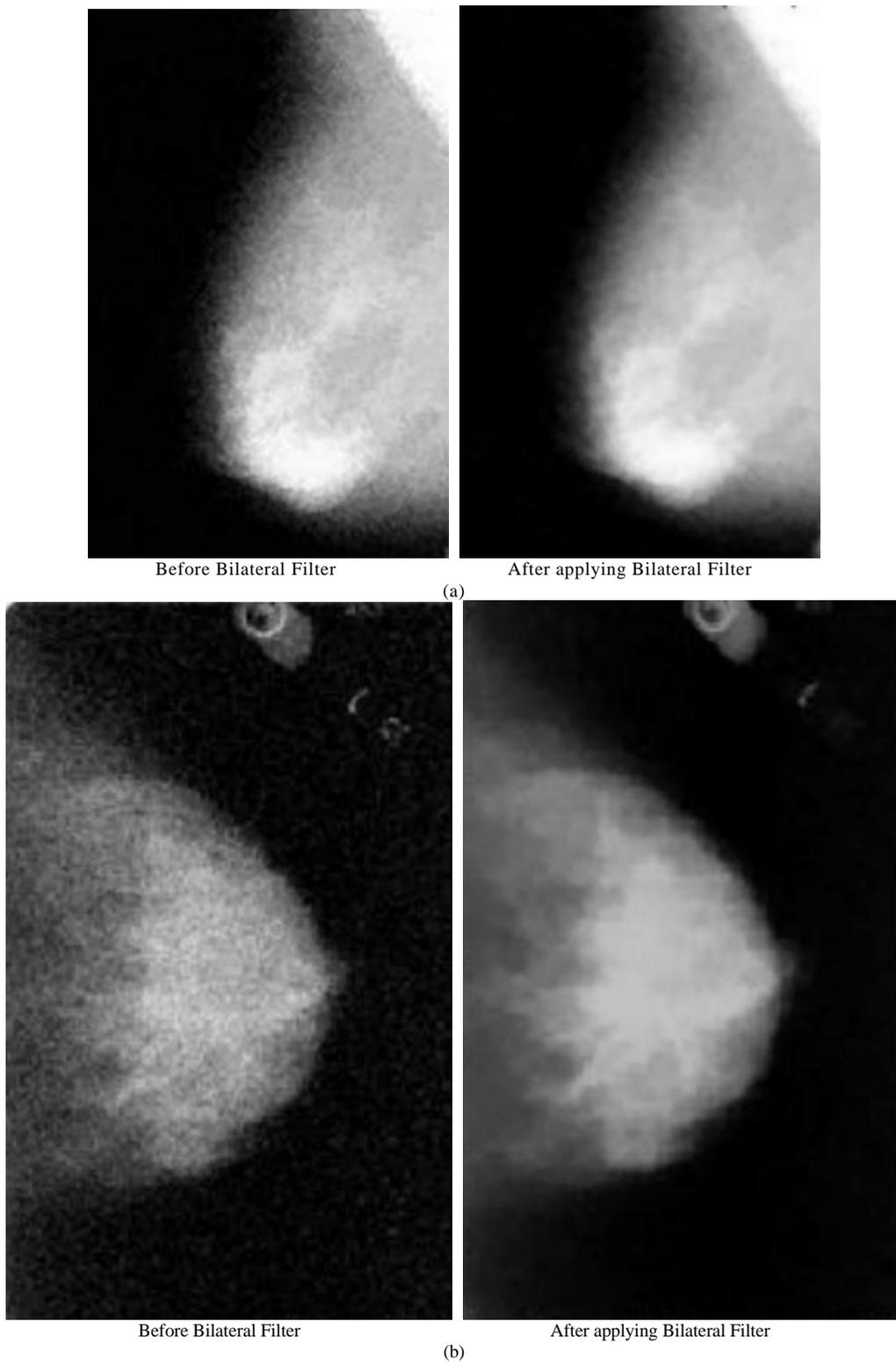
Fig. 2.   (a & b) Results of Bilateral Filter on mammogram images
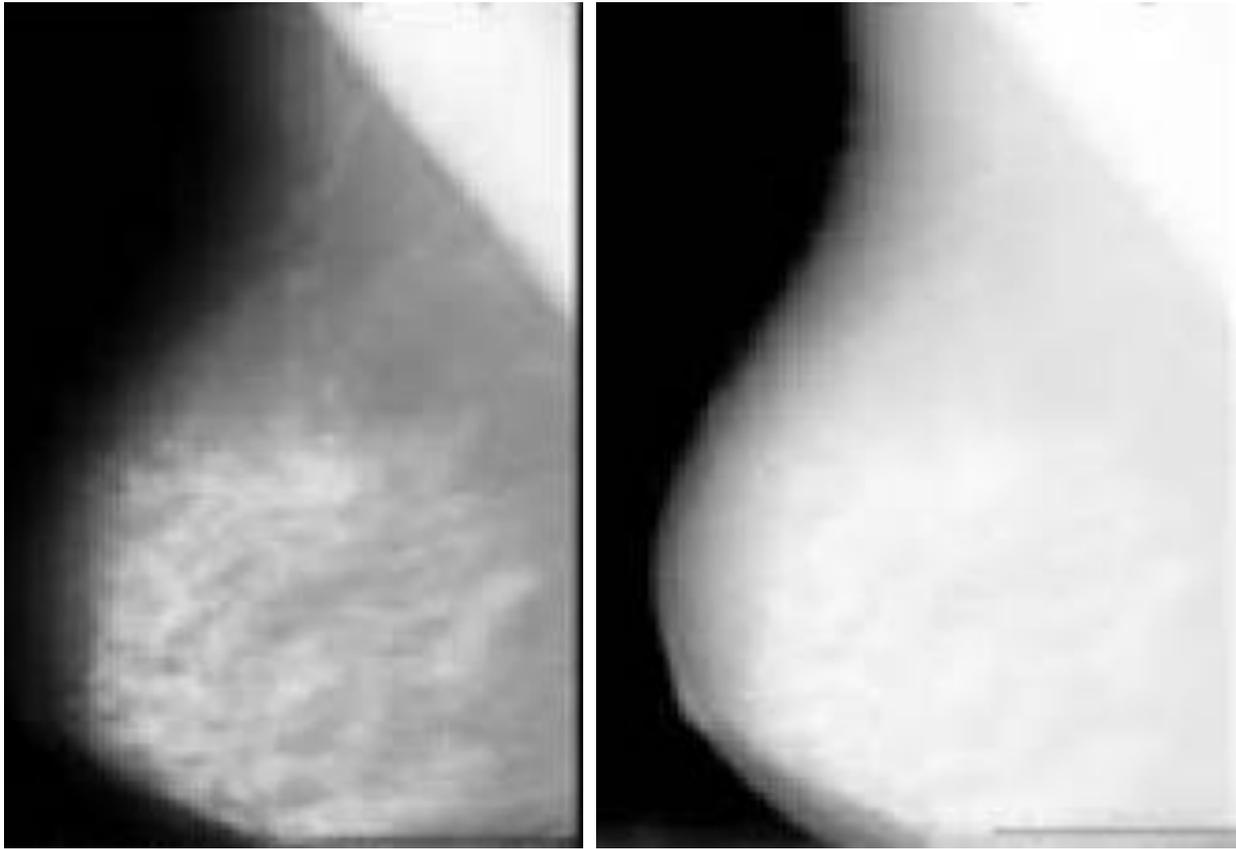
Fig. 3. Results of Log Transformation on mammogram images

Where Eint = internal energy, Eimg = forces of the image, and Econ = External constraint forces. Internation energy can be calculated by using 1st and 2nd derivatives of the parametric curve equation and it calculates by using equation (4).

$$E_{int} = E_{elastic} + E_{bending} \qquad (4)$$
$$= \int_{s} \frac{1}{2}(\alpha|v_s|^2 + \beta |v_{ss}|^2)d$$

External energy shows the image properties like edges or noise in the image. It can be calculated by using following equation (5).

$$E_{ext} = \int_{s} E_{image}(v(s))ds \qquad (5)$$

Thus to start the contour, the first initial point is selected from the white part after thresholding and then applied to the image that has been returned by applying bilateral filtered image so that actual breast part can be extracted from the original image.

This active contour process returns the breast part only that can be used for enhancement and features extraction.

B. Texture Features using Gabor Filter

Gabor filter can be used to extract texture information. The texture shows a specific pattern and mammogram images has some specific pattern that represents a specific texture and pattern. Therefore, texture is the most suitable for features extraction in the case of mammograms. So the characteristics of texture can be represented by spatial frequencies and it can also be represented by their orientations. There are different types of Gabor filter that can be applied on images to extract texture features. But in mammograms 2-D Gabor filter is most suitable due to nature of images that are in 2-D form. Gabor filter is a Gaussian kernel function and that can be modulated by a sinusoidal wave of precise frequencies and orientation. To represent the 2-D Gabor filter, fowling equations can be used:

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

$$\bar{x} = x.\cos\theta + y.\sin\theta \quad and \quad \bar{y} = x.\sin\theta + y.\cos\theta$$

Where variables x, and y are the spatial variables, σ_x and σ_y represent the scaling parameters of the filter, and W is the central frequency of the complex wave. Gabor filter bank is a combination of different Gabor filters applied at different scales, frequencies and orientation (Figure 4). It is possible to generate different filter banks with different orientation and scales. In this paper, Gabor filter bank has been created by applying two frequencies, two scales, and two orientations. For this purpose, following values has been used for generation of the filter bank. After calculating these filters, convolution is required to apply on the original images. So these eight filters are convolved with the original images so it returns eight new convolved images. After applying Gabor filter bank, there are

magnitude values of the Gabor transform. These magnitudes represent changes very slowly with displacement. After that some statistical information has been extracted from these Gabor filtered images. Mean, variance, skewness, kurtosis, entropy and energy have been calculated from all these filtered images and then make a vector for classification.
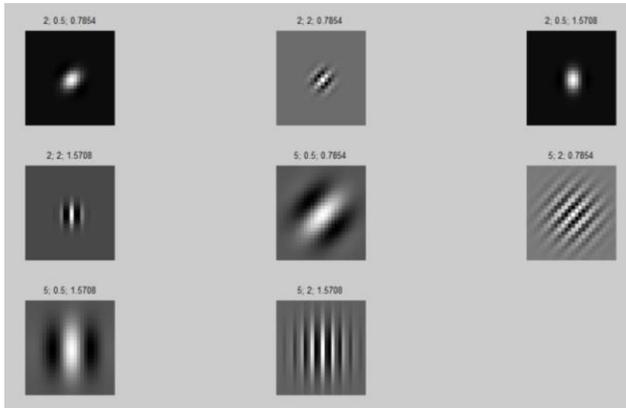
Fig. 4.   Gabor Filters with 2 orientations and 4 scales

III.   CLASSIFICATION USING ADABOOST

Classification is the process to differentiate into classes by using some characteristics. In the literature, many different classifiers are available that can classify individually. Ensemble classification used different weak classifiers and combined intelligently to combine those classifiers to improve the performance of classification. One of the most important ensemble classifiers is AdaBoost that is also known as adaptive boosting. This AdaBoost was proposed by [14] and it improves the simple classifier by using the iterative procedure. In this iterative procedure, during each iteration, there is a process to improve the misclassified samples. This procedure increased weights of misclassified patterns and decreased the weights of correctly classified samples during each iteration. In this way, weak classifiers are given more preferences and these weak classifiers are forced to learn more by using difficult samples [14]. In this way, classification performance improves during this iterative weight adjustment procedure. These adaptive weights can be used for the classification of new samples. In this way, algorithm supposed that the training set contains m samples and these samples are labelled as -1 and +1. In this way, classification of the new sample can find out by using voting for all classifiers Mt with weights αt. Mathematically, it can be written as:

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t . M_t(x))$$

Pseudocode of the AdaBoost is given in Figure 5.

Require: I (a weak inducer), T (the number of iterations), S (training set)
Ensure: $M_t, \alpha_t; t = 1, \ldots, T$
1: $t \leftarrow 1$
2: $D_1(i) \leftarrow 1/m; i = 1, \ldots, m$
3: **repeat**
4:     Build Classifier M_t using I and distribution D_t
5:     $\varepsilon_t \leftarrow \sum_{i: M_t(x_i) \neq y_i} D_t(i)$
6:     **if** $\varepsilon_t > 0.5$ **then**
7:         $T \leftarrow t - 1$
8:         exit Loop.
9:     **end if**
10:     $\alpha_t \leftarrow \frac{1}{2} \ln(\frac{1-\varepsilon_t}{\varepsilon_t})$
11:     $D_{t+1}(i) = D_t(i) \cdot e^{-\alpha_t y_t M_t(x_i)}$
12:     Normalize D_{t+1} to be a proper distribution.
13:     $t++$
14: **until** $t > T$

Fig. 5.   Pseudo code of Adaboost Classifier

IV.   RESULTS AND DISCUSSION

To test the performance of the proposed method, different quantitative measures have been used. Accuracy, sensitivity, specificity and Area under The Curve (AUC) have been used. These can be calculated by using mathematical equations shown in equations (5), (6) and (7).

$$\frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (5)$$

Sensitivity can be calculated by using

$$\frac{(TP)}{(TP + FN)} \quad (6)$$

Specificity can be calculated by using

$$\frac{(TN)}{(TN + FP)} \quad (7)$$

Where TP is True positive, FP is false positive FN is false negative and TN is true negative

I have performed three types of experiments by dividing the testing data into different ratios so that there should be no bias in training and testing. To overcome such type of problems, three different ratios like 40-60 mean 40% for training and 60% for testing, 50-50 mean 50% for training and 50% for testing and 60-40 mean 60% for training and 40% for testing has been used. We measure accuracy, sensitivity and specificity and by using these Area under the Curve (AUC) also calculated to show the performance of the proposed method. I have used different classifiers to test the performance to show that which classifier is best suitable for this problem (Figure 6). Results have been shown in Tables 2 and 3. These results show that by using proposed method with ensemble classifiers, it performs best in all cases. Support Vector Machine (SVM), K nearest neighbour (KNN), artificial Neural

Network (ANN) and ensemble classifier has been compared by using the same features set. These results show that ensemble has the best accuracy, sensitivity, specificity as well as AUC. Tables 2 and 3 shows enhancement is better to improve performance. Therefore, to compare with existing methods, I used ensemble classifier by using enhancement and also select the best ration that is 60-40 where 60% data used for training and 40% used for testing and results shown in Figure 6.

TABLE II.   COMPARISON USING ACCURACY AND SENSITIVITY

Classifier	Training-Testing	Accuracy (%)	Sensitivity (%)
SVM	40-60	93.51	87.13
	50-50	95.57	91.88
	60-40	96.69	95.01
KNN	40-60	84.47	89.02
	50-50	86.48	90.9
	60-40	91.99	95.93
ANN	40-60	94.5	91.85
	50-50	94.62	93.82
	60-40	95.39	95.34
Ensemble AdaBoost	40-60	95.58	95.34
	50-50	95.94	95.92
	60-40	96.74	98.34

TABLE III.   COMPARISON USING SPECIFICITY AND AUC

Classifier	Training-Testing	Specificity (%)	AUC
SVM	40-60	94.37	0.9773
	50-50	95.04	0.9813
	60-40	95.56	0.9844
KNN	40-60	93.5	0.9737
	50-50	94.1	0.9761
	60-40	94.47	0.9852
ANN	40-60	96.3	0.9806
	50-50	98.35	0.996
	60-40	98.9	0.9997
Ensemble AdaBoost	40-60	98.31	0.9877
	50-50	98.85	0.9956
	60-40	99.01	0.9948

After that I have compared with existing methods to test the performance of proposed method. Results have been shown in Table 2. These results show that proposed method shows best results as compared to all other existing methods in both the accuracy as well as the sensitivity. The main reason for the improved performance is good segmentation of the breast part, most suitable features extraction and ensemble classifier also plays an important role to increase the performance. Same performance measures are used as well as other parameters for classifiers.
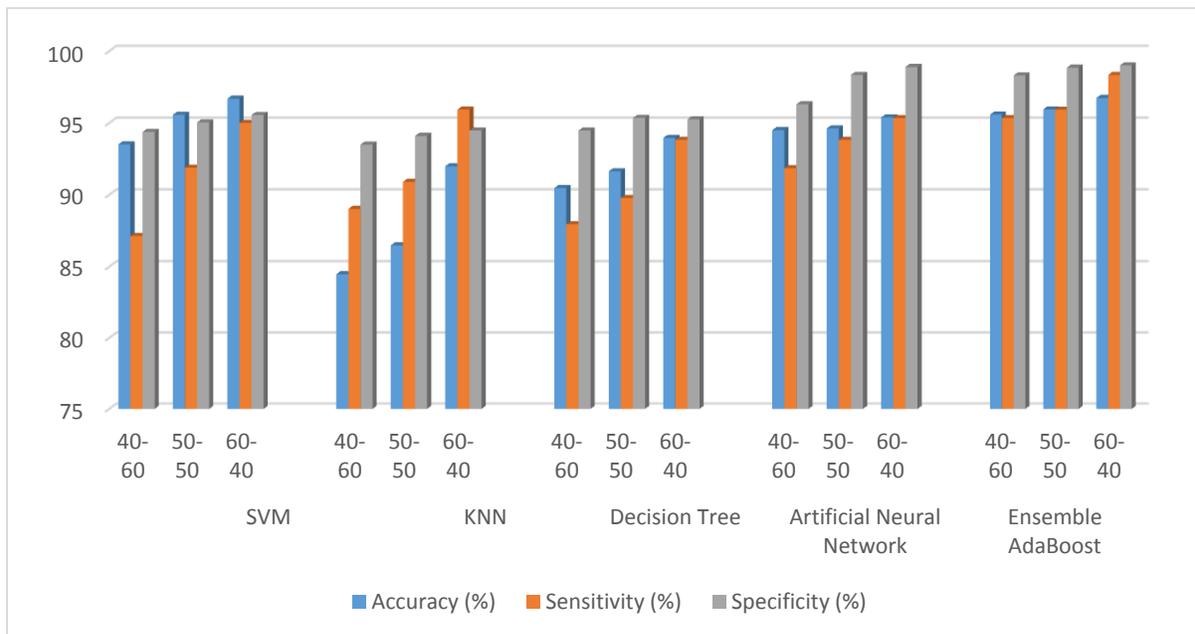
Fig. 6.   Comparison of different classifiers

V. CONCLUSION

In this paper, I have proposed a computer aided diagnosis system that performs three different tasks. In the first task, breast segmentation has been performed by using a mixture of bilateral filter, log transformation, adaptive active contour and entropy. Then enhancement has been performed by using the concept of Partitioned Iterated Function System. At the end most suitable texture features has been extracted and classified by ensemble classifier that performs well as compare to other classifiers. Due to these contributions, proposed system performs well. In the future, I will try to use some other features for classification. Deep Learning is also well suited for this problem. So in the future, deep learning concept can be applied to test the performance.

REFERENCES

[1] Timp, S., Varela, C., & Karssemeijer, N. (2007). Temporal change analysis for characterization of mass lesions in mammography. IEEE Transactions on Medical Imaging, 26(7), 945–953.

[2] Giger, M.L., Nishikawa, R.M., Kupinski, M., Bick, U., Zhang, M., Schmidt, R.A., Wolverton, D.E., Comstock, C.E., Papaioannou, J., Collins, S.A., Urbas, A.M., Vyborny, C.J., Doi, K., "Computerized detection of breast lesions in digitized mammograms and results with a clinically-implemented intelligent workstation", in Computer Assisted Radiology and Surgery, Lemke, H.U., Inamura, K., Vannier, M.W., eds., Elsevier, Berlin, Germany, pp. 325-330, 1997.

[3] Bird, R.E., Wallace, T.W., Yankaskas, B.C.:, "Analysis of cancers missed at screening mammography", Radiology , 184 , pp.613-617, 1992.

[4] Bird, R.E., Professional Quality assurance for mammography screening programs, Radiology 177, pp.587. 1990.

[5] Bellotti, R., De Carlo, F., Tangaro, S., Gargano, G., Maggipinto, G., Castellano, M., … De Nunzio, G. (2006). A completely automated CAD system for mass detection in a large mammographic database. Medical Physics, 33(8), 3066–3075.

[6] Varela, C., Tahoces, P. G., Méndez, A. J., Souto, M., & Vidal, J. J. (2007). Computerized detection of breast masses in digitized mammograms. Computers in Biology and Medicine, 37(2), 214–226.

[7] F. Sahba, A. Venetsanopoulos, Breast mass detection using bilateral filter and mean shift based clustering, in: Proceedings of the 2010 International Conference on Signal Processing and Multimedia Applications (SIGMAP), 2010, pp. 88–94.

[8] Urvashi Manikpuri, Yojana Yadav, Image Enhancement Through Logarithmic Transformation, International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Volume 1 Issue 8, 2014

[9] Wei, D., Chan, H. P., Helvie, M. a, Sahiner, B., Petrick, N., Adler, D. D., & Goodsitt, M. M. (1995). Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis. Medical Physics.

[10] Kass, M.; Witkin, A.; Terzopoulos, D. (1988). "Snakes: Active contour models" (PDF). International Journal of Computer Vision. 1 (4): 321

[11] T.L. Economopoulos, P.A. Asvestas, G.K. Matsopoulos, Contrast enhancement of images using Partitioned Iterated Function Systems, Image and Vision Computing Volume 28, Issue 1, January 2010, Pages 45–54

[12] Freund Y, Schapire RE, Experiments with a new boosting algorithm. In: Machine learning: proceedings of the thirteenth international conference, 1996, pp 325–332

[13] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Transactions on Systems, Man and

[14] Cybernetics, vol. 3, no. 6, pp. 610–621, 1973

[15] Lior Rokach, Ensemble-based classifiers, Artificial Intelligence Reviews (2010) 33:1–39

[16] http://skye.icr.ac.uk/miasdb/miasdb.html.

[17] https://radiopaedia.org/cases/labelled-normal-mammograms