Proceedings of the 10th INDIACom; INDIACom-2016; IEEE Conference ID: 37465
2016 3rd International Conference on "Computing for Sustainable Global Development", 16th - 18th March, 2016
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Classification of Breast Lesions based on Laws' Feature Extraction Techniques

Sahil Bhusri
Jaypee University of Information Technology
Shimla , India
Email Id:sahilbhusri1@gmail.com

Shruti Jain
Jaypee University of Information Technology
Shimla , India
Email Id:shruti.jain@juit.ac.in

Jitendra Virmani
Thapar University Patiala
Punjab ,India
Email Id:jitendra.virmani@gmail.com

***Abstract*—Breast lesions are characterized into three classes which include primary benign, primary malignant and secondary malignant. In the present work Laws' mask texture features are computed from the ultrasound images of the breast lesions. These Laws' masks of various resolutions i.e., of length 3, 5, 7 and 9 have been used to extract the statistical features (Mean, Standard Deviation, Kurtosis, Skewness and Energy) from Laws' texture images. In the present work using the SVM classifier, an overall classification accuracy of 88.3% and the individual classification accuracy values of 95.2% , 88.6% and 91.6% have been obtained for primary benign, primary malignant and secondary malignant classes respectively.**

Keywords — Breast cancer, Classification, Laws' Feature Extraction, Primary Benign, Primary Malignant, Secondary Malignant , Ultrasound

I. INTRODUCTION

Breast Cancer being the biocide form of the cancer, in women after the Lung cancer , is the most generic form of cancer resulting in the loss of life. According to the Survey figures by American Cancer Society, Surveillance and Health Services Research in United States, total number of breast cancer cases estimated to occur in 2015 are 2,92,130 and total no of deaths estimated are 40,290 [1]. The cancer sets about in the inner lining of ducts or lobules that supply the ducts with milk [2]. Early detection of the disease doesn't only help in the proper diagnosis but also minimizes the risk of the unwanted result of the disease (death). Various early detecting techniques are available these days that include breast exam by physician, X-Ray, Ultrasonography, Magnetic Resonance Imagining (MRI) and Biopsy. In biopsy, a sample of lesion is taken out for the analysis that results in unbearable pain to patient. To help the patient and reduce unnecessary biopsies, the most frequent method includes Mammography, Ultrasonography. X-ray Mammography does not divulge soft tissues and has a very low sensitivity for the dense breasts [3], where as the Ultrasound imaging provides the non-radioactive, non invasive, real time display, low cost and better penetration ability as compared to the X-ray Mammography [4]. Ultrasonography doesn't only discriminate between the cysts and the solid lesions, but modern ultrasound also helps to distinguish h between the benign and malignant types. Benign tumors have well defined contour with round and smooth shape and margin where as the invasive type of cancer, i.e. the Malignant type of tumors are having the architectural distortions, spiculations, angular margins, irregular shape, acoustic shadowing, duct extension [5].

Primary Benign case is Fibroadenoma and Primary Malignant case is Carcinoma [1]. The secondary case of Malignancy is Metastasis. Till now, most of the work is done to distinguish between only the Benign and Malignant class but in the present work, we will distinguish between the primary and secondary occurring classes of benign and malignant cases.

II. METHODOLOGY

The experimental flow of the system follows a sequence as shown in Figure 1

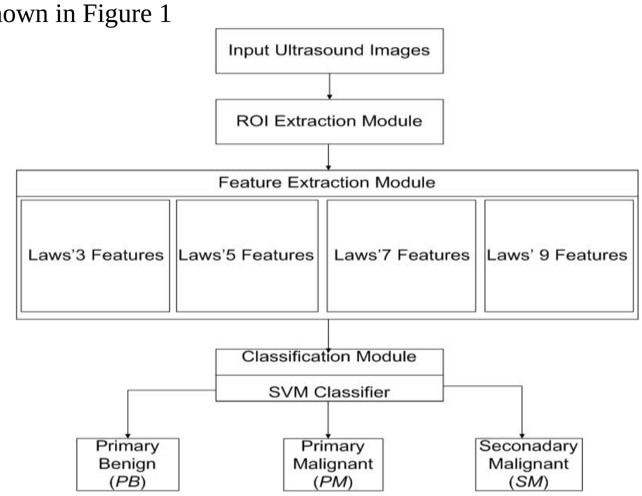


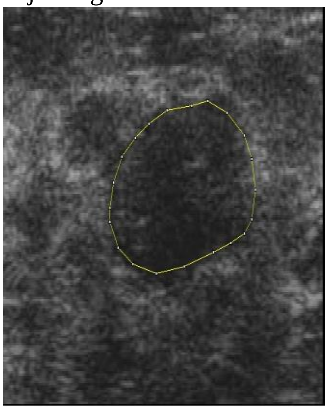
Fig. 1. Overview of system

A). Database Ultrasound Images

For analysis, the benchmark ultrasound data is taken online from [6] that is for the Primary Benign, Primary Malignant and Secondary Malignant cases. The dataset contains a total of 172 cases having 56 cases of Fibroadenoma (Primary Benign) and 74 cases of Carcinoma (Primarily Malignant) and 42 cases of Metastasis (Secondarily Malignant) and the cases of lesions during biopsy implementation are discarded.

B). ROI (Region of Interest) Extraction Module

The abnormality in the ultrasound is identified and marked with the help of an experienced radiologist and segmented with the help of software *Image J* [7].This software helps to load the image, mark the infected area and segment it. Further the segmented region is enclosed into a rectangular bounding box adjoining the boundaries of abnormality as shown in Fig [2-4]



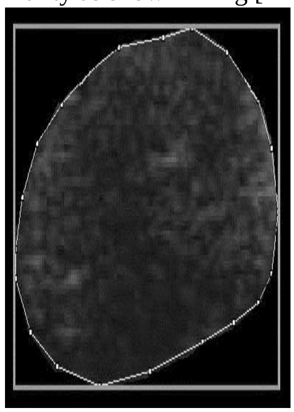


Fig. 2.1 Fig. 2.2

Figure 2.1 : ROI marked in Fibroadenoma
Figure 2.2 : Bounding Box enclosing ROI of Fibroadenoma

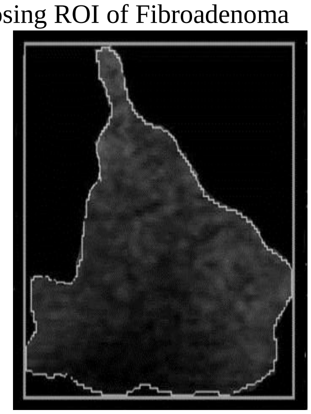
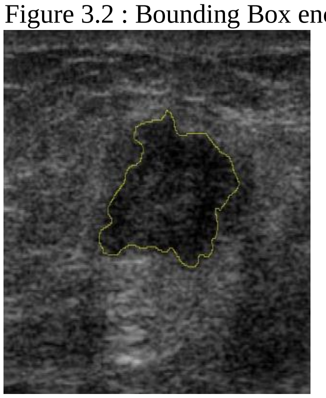


Fig. 3.1 Fig. 3.2

Figure 3.1 : ROI marked in Carcinoma
Figure 3.2 : Bounding Box enclosing ROI of Carcinoma



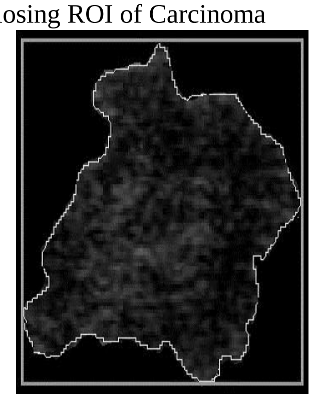


Fig.4.1 Fig.4.2

Figure 4.1 : ROI marked in Metastasis
Figure 4.2 : Bounding Box enclosing ROI of Metastasis

C). Feature extraction Module

When a region suffers from damage through any disease then that region is known as lesion, Feature extraction module extracts features by Morphological methods and Texture methods. Texture Methods are classified as 1) Signal Processing Based Methods 2) Statistical Methods and 3) Transform Domain Methods. In the present work, characterization of lesions is implemented using the Signal Processing based methods.

Signal Processing Method

A set of coherent masks known as Texture Energy Masks are used in Laws' Based Textures Features. The texture features are estimated by convolving the images with these masks. The filters are made up by the combination of two or more one dimension kernel vectors. The dimensions (*d*) of these vectors can take value d=3, 5, 7 and 9. The properties of the texture are determined by filters by performing level detection (*L*), edge detection (*E*), ripple detection (*R*), spot detection (*S*) and wave detection (*W*) [8-14]. The Laws' mask of length 3 is described as L3= [1, 2, 1], E3= [-1, 0,1], S3=[-1, 2, -1] and of dimension 5 is L5= [1, 4, 6, 4, 1], E5= [-1, -2, 0, 2, 1] , S5= [-1, 0, 2, 0,-1] , W5=[-1, 2, 0, -2, 1] and R5= [1, -4, 6, -4, 1] . The masks of different dimension are used to extract different features. The mask for different dimensions d= 3, 5 are shown in Figure 5 and Figure 6

L3L3	E3L3	S3L3
L3E3	E3E3	S3E3
L3S3	E3S3	S3S3

Fig.5 : For d=3

In Laws' Mask of dimension 3 as in figure 5, there are 9 two dimension masks which include 3 masks of identical pairs, so the number of rotational invariant texture images will be 6. There are 5 descriptors derived from ROI, so features vector length will be 6×5= 30.

L5L5	E5L5	S5L5	R5L5	W5L5
L5E5	E5E5	S5E5	R5E5	W5E5
L5S5	E5S5	S5S5	R5S5	W5S5
L5R5	E5R5	S5R5	R5R5	W5R5
L5W5	E5W5	S5W5	R5W5	W5W5

Fig.6 : for d= 5

Here in Laws' mask of dimension 5 as in Figure 6, there are 25 two dimension masks in which rotational invariant texture images of identical pairing are 10, so total invariant texture images will be 15. 5 descriptors are derived from ROI, so the feature vector length or number of features extracted will be 15 ×5= 75.

The Laws' Mask filter of length 7 will be same as Laws' mask of dimension 3and the Laws' Mask filter of length 9 will be same as of Laws' mask of dimension 5 and will have same feature vector length as dimension 3 and 5 respectively. In the analysis of Laws' Mask, there is a sequence to be followed. As an example consider the Laws' mask of length 5.

a) The Texture Image (*TI*) is obtained by convolving the input image I (i, *j*) with the 2-D mask

$$\mathrm{TI_{E5E5}} = \mathrm{I_{i,j}} \otimes \mathrm{E5E5} \tag{1}$$

b) The contrast of the texture image obtained from equation (1) is normalized

$$\mathrm{Normalize\ (TI_{mask})} = \frac{\mathrm{TI_{mask}}}{\mathrm{TI_{L5L5}}} \tag{2}$$

c) The Texture Energy Measurement (TEM) filters are used to pass the Texture image

$$\mathrm{TEM_{i,j}} = \sum_{u=-5}^{5}\sum_{v=-5}^{5} \mathrm{Normalize(TI_{i+u,j+v})} \tag{3}$$

d) To obtain 15 rotationally invariant TEM's that are denoted as *TR* are obtained by combining the 25 TEM descriptors

$$\mathrm{TR_{E5L5}} = \frac{\mathrm{TEM_{E5L5}} + \mathrm{TEM_{L5E5}}}{2} \tag{4}$$

e) Five statistical parameters are determined, that are the Mean, Standard Deviation, Skewness, Kurtosis, Entropy. Here M × N is the dimension of the image

1) Mean (m): It describes the mean intensity value with in texture image.

$$Mean = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(TR_{i,j})}{M \times N} \tag{5}$$

2) Standard Deviation (SD): It is used to measure the variability.

$$\mathrm{SD} = \sqrt{\frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(\mathrm{TR_{i,j}} - \mathrm{Mean})^2}{\mathrm{M \times N}}} \tag{6}$$

3) Skewness: It measures of the asymmetry of the probability distribution of a random variable that is real valued.

$$\mathrm{Skewness} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(\mathrm{TR_{i,j}} - \mathrm{Mean})^3}{\mathrm{M \times N \times SD}} \tag{7}$$

4) Kurtosis: It measures of the probability distribution's shape of a random variable that is real valued.

$$\mathrm{Kurtosis} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(\mathrm{TR_{i,j}} - \mathrm{Mean})^4}{\mathrm{M \times N \times SD^4}} - 3 \tag{8}$$

5) Entropy: It measures the randomness of the elements of the image.

$$\mathrm{Entropy} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(\mathrm{TR_{i,j}})^2}{\mathrm{M \times N}} \tag{9}$$

D). Classification Module

The technique of the clustering the testing samples into distinct classes refers to classification. Classification is characterized into two types: supervised classification and the unsupervised classification. If the classes are already defined for the training sets then classification is supervised classification else it is unsupervised classification. In the present work, SVM classifier is used for the classification task. The SVM classifier is included in class of supervised classification. For the implementation of SVM classifier, LibSVM library has been used [16]. SVM works on the fundamental concept of decision planes in which the decision boundaries are defined. In kernel based classifiers, nonlinear mapping of training data from input space to higher dimensional feature space is done using the kernel functions. For the classification task, Gaussian radial basis function kernel's performance is explored. The choice of the regularization parameter C and kernel parameter γ is always a decisive step for having a good generalization performance. By doing the extensive search, that is carried out in the parameter space for the values of $C \in \{2^{-4}, 2^{-3}, \dots 2^{15}\}$ and $\gamma \in \{2^{-12}, 2^{-11}, \dots 2^{4}\}$ the optimal values for C and γ are obtained [15, 17-23].

III. Results and Discussion

The results of classification computed with Laws' Mask of different dimensions 3, 5, 7, 9 are tabulated in Table 1-4.

Table-1 Result of SVM classifier with Laws' Mask of dimension 3 having TFV length 30

TFV(*l*)	CM				ICA	OCA
		PB	PM	SM		
TFV(30)	PB	20	0	1	95.2 %	85.7%
	PM	1	37	6	84.0 %	
	SM	1	2	9	75.0 %	

Note: TFV: Texture Feature Vector , CM: Confusion Matrix, OCA: Overall classification Accuracy, ICA: Individual Classification Accuracy, PB: Primary Benign, PM: Primary Malignant, SM: Secondary Malignant, *l*: length of TFV

Table-2 Result of SVM classifier with Laws' Mask of dimension 5 having TFV length 75

TFV (*l*)	CM				ICA	OCA
		PB	PM	SM		
TFV(75)	PB	18	0	3	85.7 %	88.3%
	PM	0	39	5	88.6 %	
	SM	0	1	11	91.6 %	

Note: TFV: Texture Feature Vector, CM: Confusion Matrix, OCA: Overall Classification Accuracy, ICA: Individual Classification Accuracy, PB: Primary Benign, PM: Primary Malignant, SM: Secondary Malignant, *l*: length of TFV

Table-3 Result of SVM classifier with Laws' Mask of dimension 7 having TFV length 30

TFV(*l*)	CM				ICA	OCA
		PB	PM	SM		
TFV(30)	PB	11	0	10	52.3 %	051.9 %
	PM	1	23	20	52.2 %	
	SM	5	1	6	50.0 %	

Note: TFV: Texture Feature Vector, CM: Confusion Matrix, OCA: Overall Classification Accuracy, ICA: Individual Classification Accuracy, PB: Primary Benign, PM: Primary Malignant, SM: Secondary Malignant, *l*: length of TFV

Table-4 Result of SVM classifier with Laws' Mask of dimension 9 having TFV length 75

TFV(*l*)	CM				ICA	OCA
		PB	PM	SM		
TFV(75)	PB	17	0	4	80.9 %	83.1 %
	PM	0	37	7	84.0 %	
	SM	1	1	10	83.3 %	

Note: TFV: Texture Feature Vector, CM: Confusion Matrix, OCA: Overall Classification Accuracy, ICA: Individual Classification Accuracy, PB: Primary Benign, PM: Primary Malignant, SM: Secondary Malignant, *l*: length of TFV

In this experiment, the classification performance of different TFVs derived using Laws' masks of different lengths is calculated with SVM classifier. The best overall classification accuracy of the system is 88.3% using Laws' mask of length 5. The best individual classification accuracy (ICA) for primary benign is 95.2 % using Laws' mask of length 3. Similarly, best ICA for primary malignant case is 88.6 % using Laws' mask of length 5 and for secondary malignant case best ICA achieved is 91.6 % using Laws' mask of length 5.

IV. Conclusion and Future Work

The best OCA achieved using the proposed CDA system is 88.3% and is coming out by texture feature vector having length 75 with dimension 5. But the cases with the dimension 7 are misclassified, so in future authors will concentrate to increase the performance of the system with dimension 7 and also look to increase the overall performance of the whole system.

V. Acknowledgment

The authors wish to acknowledge Dr Shruti Thakur, Indira Gandhi Medical College, Shimla, Himachal Pradesh, India for her constant patronage and support in carrying out this research work.

References

[1] Breast Cancer facts and figures 2015-16, *American Cancer Society*, [online] 2015 http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-046381.pdf
[2] What is cancer? *MNT Knowledge Center*, [online], http://www.medicalnewstoday.com/articles/37136.php

[3] H.M. Zonderland .”The role of ultrasound in diagnosis of breast cancer,” *Semin Ultrasound CT MR* . ,vol. 21 , pp. 317-324 , 2000

[4] Minavathi , Murali .S , M.S.Dinesh , “ Classification of Mass in breast Ultrasound Images Using Image Processing Techniques ” International Journal of Computer Applications , vol. 42, issue 10 ,pp. 29-36 , 2012

[5] MO Obajimi, OO Akute, AO Afolabi, AA Adenipekun, AO Oluwasola, EEU Akang, RU Joel, ATS Adeniji-Sofoluwe, O Funmi, G Newstead, R Schmitt, C Sennet ,’’BI-RADS Lexicon : An Urgent Call for the standardization of Breast Ultrasound in Nigeria” *Annals of Ibadan Postgraduate Medicine* Vol. 3 (1) : pp. 82-88 , 2005

[6] *Ultrasound Cases.info* , [online] ,2015 http://www.ultrasoundcases.info/category.aspx?cat=67

[7] *Image Processing and Analysis in JAVA* ,Image J 1.49 vesion1.6.024 *http://imagej.nih.gov/ij/download/win32/ij149-jre6-64.zip*

[8] K.I. Laws, “Rapid texture identification”, in *Proceedings of SPIE Image Processing for Missile Guidance*, 1980, pp. 376-380.

[9] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, “Characterization of primary and secondary malignant liver lesions from B-mode ultrasound”, *Journal of Digital Imaging*, vol. 26, pp. 1058-1070, 2008.

[10] M. Rachidi, A. Marchadier, C. Gadois, E. Lespessailles, C. Chappard and C. L. Benhamou, “Laws' masks descriptors applied to bone texture analysis: an innovative and discriminant tool in osteoporosis”, *Skeletal Radiology*, vol. 37, pp. 541-548, 2008.

[11] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, “Prediction of cirrhosis from liver ultrasound B-mode images based on Laws' mask analysis”, in *Proceedings of IEEE International Conference on Image Information Processing*, *ICIIP-2011*. Himachal Pradesh, India, 2011, pp. 1-5.

[12] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, “Neural network ensemble based CAD system for focal liver lesions from B-mode ultrasound”, *Journal of Digital Imaging*, vol. 27, pp. 520-537, 2014.

[13] G.H. Seng, H.Y. Chai and T.T. Swee, “Research on Laws' mask texture analysis system reliability”, *Reasearch Journal of Applied Sciences, Engineering and Technology*, vol. 7, pp. 4002-4007, 2014.

[14] Kriti and J. Virmani, “Breast density classification using Laws' mask texture features”, *International Journal of Biomedical Engineering and Technology*, vol. 19, no. 3, pp. 279-302, 2015.

[15] Kriti, J. Virmani, N. Dey and V. Kumar, “PCA-PNN and PCA-SVM based CAD systems for breast density classification”, in A.E. Hassanien et al. (eds.) *Applications of Intelligent Optimization in Biology and Medicine: Current Trends and Open Problems*, vol. 96, pp. 159-180, 2015.

[16] C.C. Chang and C. J. Lin, “LIBSVM, A library of support vector machines”.

[17] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, “SVM based characterization of liver cirrhosis by singular value decomposition of GLCM matrix”, *International Journal of Artificial Intelligence and Soft Computing*, vol. 3, pp. 276-296, 2013.

[18] A.E. Hassanien, N.E. Bendary, M. Kudelka and V. Snasel, “Breast cancer detection and classification using support vector machines and pulse coupled neural network”, in *Proceedings of 3rd International Conference on Intelligent Human Computer Interaction IHCI 2011*, Prague, Czech Republic, 2011, pp. 269-279.

[19] A.T.Azar and S.A. El-Said, “Performance analysis of support vector machine classifiers in breast cancer mammography recognition”, *Neural Computing and Applications*, vol. 24, pp. 1163-1177, 2014

[20] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, “PCA-SVM based CAD system for focal liver lesion using B-mode ultrasound images”, *Defence Science Journal*, vol. 63, pp. 478-486, 2013.

[21] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, “Prediction of liver cirrhosis based on multiresolution texture descriptors from B–mode ultrasound”, *International Journal of Convergence Computing*, vol. 1, pp. 19-37, 2013.

[22] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, “SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors”, *Journal of Digital Imaging*, vol. 26, no. 3, pp. 530-543, 2012.

[23] J. Virmani and Kriti, “Breast tissue density classification using wavelet based texture descriptors”, in *Proceedings of 2nd International Conference on Computer and Communication Technologies (IC3T 2015)*, 24-26 July, 2015, Hyderabad, vol. 381, pp. 539-546, 2015.