Int J Pharm Bio Sci 2016 April; 7(2): (B) 617 - 624

Original Research Article **Biotechnology**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

BREAST LESIONS CLASSFICATION USING THE AMALAGATION OF MORPHOLOGICAL AND TEXTURE FEATURES

SAHIL BHUSRI*[1], SHRUTI JAIN[1] AND JITENDRA VIRMANI[2]

[1]*Jaypee University of Information Technology, Solan, HP, India*
[2]*Thapar University, Patiala ,Punjab ,India.*

ABSTRACT

The aim of this paper is to classify the breast lesions using the combination of two feature extraction techniques i.e. morphological features and the texture features. The breast lesions are characterized into two categories Benign and Malignant. Morphological Features computes Area, Perimeter, Convex area, Diameter , Major axis , Minor axis, Extent , Eccentricity, Euler no ,Solidity and Orientation where texture feature /are computed using the statistical features using FOS, GLCM, GLRL, Edge, GLDS, SFM,NGTDM, based statistical feature extraction methods. SVM classifier is extensively used for classification. Using the combination of morphological features and statistical features, the overall classification accuracy of 83.1 % is achieved and the combination of morphological and first order statistics yields the classification accuracy of 89.6%.

KEYWORDS: Breast cancer, Morphological features, Statistical features, Ultrasound

SAHIL BHUSRI
Jaypee University of Information Technology, Solan, HP, India.

***Corresponding author**

INTRODUCTION

Breast cancer is the second most fatal disease after lung cancer in women. Being the biocide form of cancer that usually starts off in the lobules supplying milk to the ducts or in the ducts that carry the milk.[1] To curtail the undesired outcomes of this fatal disease, various screening methods available that doesn't only helps to diagnose the disease but can be useful if detected in the early stage. The various screening tests include breast exam by physician (in initial stage), X-ray, Ultrasonography, Magnetic Resonance Imagining (MRI). Biopsy is also a way in which a sample of tissue from apparent abnormality is taken out for the analysis resulting unbearable pain to patient. To reduce redundant biopsies, the most frequent methods prescribed are Mammography, Ultrasonography. Imaging by Ultrasound offers non-radioactive, non invasive, real time display, low cost and better penetration ability as compared to the X-ray Mammography.[2]. Benign tumors have an oval shape thin and consistent capsule where the invasive type of cancer, Malignant type tumors don't have consistent capsule , irregular in shape , have shape and duct extension.[3] To make it more clarify and identifiable computer aided diagnosis systems were introduced. These CAD tools don't only help doctor but also decrease the rate of biopsies. To differentiate between benign type and the malignant type Alveranga et al.[4] has explored seven shape based features on the ultrasound images. Linear discriminant analysis was used to find five features. By using the circularity and convex polygon technique, Normalized radial value and the overlap ratio was calculated. The accuracy achieved in this case is 83 %. To distinguish between the benign and the malignant type, Alveranga et al. [5] uses the combination of morphological features with the texture features on database set of 246 ultrasound images and accuracy with the fisher linear discriminant analysis (FLDA) classification achieved was 85.3 %. Wu et al.[6] concluded an accuracy of 92.86% on 210 ultrasound images using SVM classification using the combination of auto covariance texture features and the morphological features .Different types of feature extraction tools and classifiers can be used to characterize the breast lesions. In this work, we will characterize between the benign and malignant classes using the combination of the morphological features with the texture feature extraction technique with SVM classifier and the main advantage of taking the combination of these features is that morphological feature alone provides only geometrical information of lesions and the statistical features about the texture . The combination of both these will be used will be used to extract the information about the shape and texture of the lesions which will increase the accuracy of detection.

METHODOLGY

In this work, sequences of steps are followed to characterize the breast lesions. The experimental flow of the system is shown in Figure 1.

A). Database Ultrasound Images

The ultrasound data used, is available online[7]. The dataset contains total no of 167 cases having 51 cases of Benign and 121 cases of Malignant class.

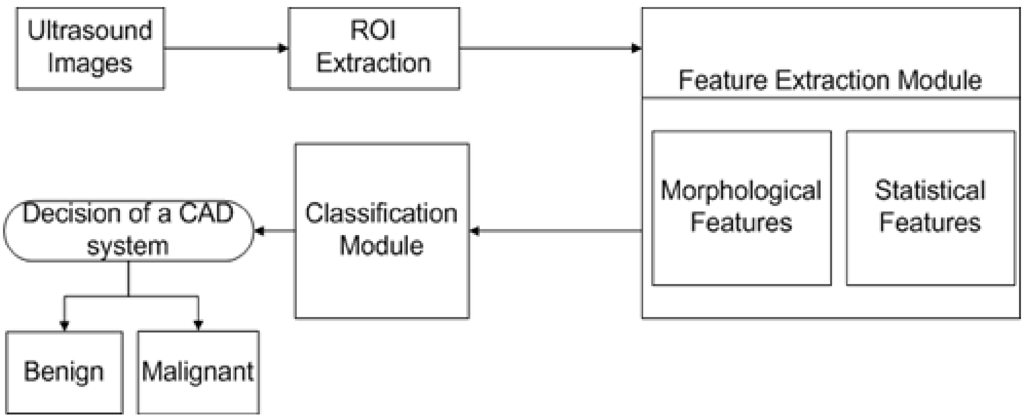

Figure 1
Experimental Workflow

B). ROI (Region of Interest) Extraction Module

To make correct decisions in CAD systems, it is very essential to identify the correct region containing the abnormality .With the guidance of an experienced radiologist, the deformity in the breast ultrasound is identified and then segmented with the help of software *Image J*.[8] This software supports to mark the infected area and segment it. A rectangular bounding box is used to enclose the deformity by adjoining the boundaries of abnormality as shown in Fig [2-3]

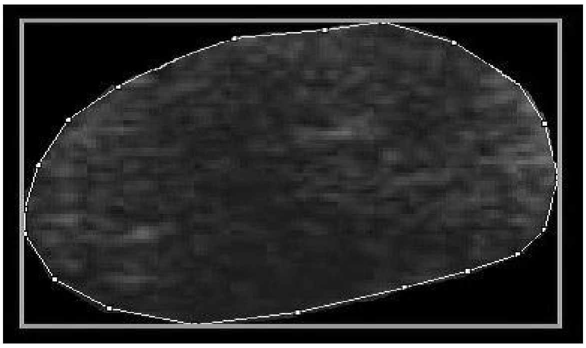

Figure 2
Bounding Box enclosing ROI of Benign Case

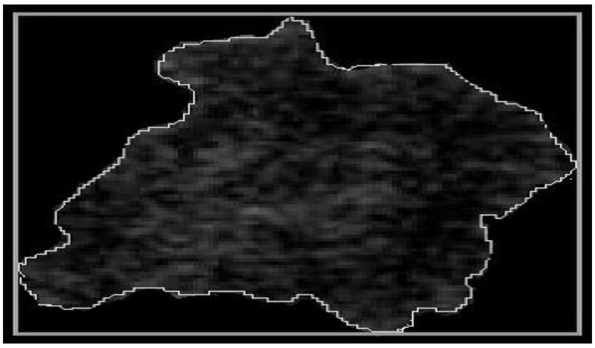

Figure 3
Bounding Box enclosing ROI of Malignant case

C). Feature extraction module

Lesion is a region that endures from damage through any disease, various types of mathematical features have been extracted *i.e. morphological features and Texture features.* Morphological features are used to extract shape based features whereas the texture features define the property of the surface; these are further categorized into three classes i.e., (a) Signal processing based methods (b) Statistical methods(c) Transform domain methods. Signal processing based methods uses Laws' Mask of different dimensions 3, 5, 7 and 9 to extract features. Statistical features includes the first order statistics, Second order statistics, Higher order statistics etc. whereas the transform domain methods includes the Fourier power spectrum ,Wavelet based methods, Gabor Features and FPS features. In this work we characterize the lesions using the combination of the Morphological features and Statistical features.

1. ***Morphological Methods***

Morphological methods include the shape based properties which includes Area, Perimeter, Convexity, Eccentricity, Extent, Hole Area Ratio (HAR) , Euler No and Solidity are calculated over the entire class of Benign and Malignant.[9]

a. Area: It calculates the area of the lesion.
b. Perimeter: It calculates the perimeter of the lesion
c. Convexity: It is the ratio of the perimeter of the convex hull to the overall contour.
d. Diameter: It is the diameter of the circle which have an equivalent area as the region
e. Major Axis and Minor Axis: These are the diameters of the ellipse, where major axis is the longest diameter and the minor axis is the smallest diameter.
f. Eccentricity: It is the ratio of the minor axis to the major axis. Its value always lies between the 0 and 1.
g. Extent: It is the ratio of the pixels in the bounding box area to the pixels present in the region.

$$\text{Extent} = \text{Area} / \text{Bounding Box} \quad (1)$$

h. Solidity: It gives the extent to which the given shape is convex or concave.

$$\text{Solidity} = \text{Area} / \text{Convex Area} \quad (2)$$

i. Euler No: It defines the relationship between the no of contiguous part and the no of holes in the shape.
j. Orientation: It tells about the angular position of the region how it is placed.

2. ***Texture Based Features***
An ultrasound image has pixels having different gray level intensities. Statistical features can be computed using first order statistics, second order statistics and higher order statistics based on the distribution of gray level intensities in a image.

a) First order statistics
These are the approaches that uses image histogram's moments to describe the texture. They compute the average gray level, Randomness, Roughness, Uniformity and entropy.[10] Let Z_i be the random variable denoting the gray levels and $p(z_i)$ denotes the values on corresponding histogram.

i) Mean (*m*)*:* It is the mean intensity value with in texture image.

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \tag{3}$$

ii) Randomness (*R*)*:* It checks whether the surface is rough or smooth depending on the values of variance where Variance (σ) tells about the range of intensities present in the image

$$R = 1 - \frac{1}{1+\sigma^2(z)} \tag{4}$$

iii) Entropy (*e*)*:* It is defined to measure the randomness of the elements of the image.

$$e = -\sum_{i=0}^{L-1} p(z_i)\log_2 p(z_i) \tag{5}$$

iv) Uniformity (*U*)*:* It is used to measure the uniformity of an image in the range [0,1].

$$U = \sum_{i=0}^{L-1} p^2(z_i) \tag{6}$$

v) N^{th} Order Moment ($\mu_n(z)$)*:* It is used to describe the moments of different order present in an image.

$$\mu_n(z) = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i) \tag{7}$$

Here in *equation (7),* by having value of *n=2 ,* variance can be calculated.

b) Second order statistics
Second order statistics includes the computations with the GLCM (Gray level co-occurrence matrix). GLCM tells how frequent combinations of pixels pairs having different gray level occurs in an image having separation of different dimensions in different directions say 0°, 45°, 90°, 135°. Total of 13 GLCM features are taken for computation in this paper , these features includes contrast ,entropy, sum entropy, difference entropy , correlation, inverse difference moment, sum average, , variance, sum variance difference variance information measures of correlation-1 and 2 , Angular second moment.[11-13]

c) Higher order statistics
Higher order statistics are computed with the use of GLRLM (Gray Level Run Length Matrix) .Texture features are computed using the different combinations of intensities at relative position of each other. Gray level run is made up by the set of consecutive pixels of gray levels that are collinear to each other and run length is the no of times a run occurs. 11 GLRLM features that are computed in this work are long run emphasis, short run emphasis, low gray level run emphasis, high gray level run emphasis, run length non-uniformity and run percentage, short run low gray level emphasis ,long run low gray level emphasis, gray level non uniformity, short run high gray level emphasis, long run high gray level emphasis.[14-15]

d) Other Statistical Features

i) GLDS Features

GLDS (Gray Level Difference Statistics) computes contrast, energy, entropy homogeneity, and mean on the basis of the co-occurrence of a pixel pair that have difference in gray levels separated by a particular distance .[16-17]

ii) EGDE Features

There is always more information present in the edges than the other parts. The spatial variation in an image is computed by calculating the gradient value. If there are abrupt changes present then gradient will be high else it will be low. Edge features computes two features: Absolute gradient mean and Absolute gradient variance.

iii) SFM

SFM (Statistical Feature Matrix) computes coarseness, contrast, periodicity and roughness of pixels at different distances within an image.[18]

iv) NGTDM Features

NGTDM (Neighborhood Gray Tone Difference Matrix) computes busyness, coarseness, complexity, contrast, and strength and considers a difference between the gray level between pixels.[19]

D). Classification Module

The classification is the process of clustering the testing samples into the corresponding classes. Classification is characterized into two types supervised classification and the unsupervised classification. Classification is supervised classification if the classes are already defined for the training sets otherwise it is unsupervised classification. Different types of classifier available are k-NN ,PNN ,SVM etc. k-NN and PNN are conceptually of similar types , both requires more time for computation because of more no of iterations and more no samples, so we prefer SVM. The SVM classifier is included in class of supervised classification. LibSVM library has been used for the implementation of SVM classifier.[20] SVM is working on the elementary approach of decision planes in which the decision boundaries are defined. In kernel based classifiers, the kernel functions are responsible for the nonlinear mapping from input space to higher dimensional feature space of the training data. For the classification task, Gaussian radial basis function kernel's performance is explored. For having a good performance the choice of the regularization parameter C and kernel parameter γ is always a decisive step.[21-25]

RESULTS AND DISCUSSION

The classification results computed with the morphological and statistical features are tabulated in given Table 1 to Table 4.The classification performance obtained by morphological features (MFV) and all the statistical features (SFV) is shown in table 1.

Table 1

Result of SVM classification of Morphological (11) and all Statistical features (46)

FV	l	CM			OCA (%)	ICA_M(%)	ICA_B(%)
MFV	11		B	M	81.1	71.4	85.7
		B	15	6			
		M	8	48			
SFV	46	B	12	9	79.2	57.1	85.7
		M	8	48			

Note:*FV: Feature Vector, MFV: Morphological feature vector, SFV: Statistical feature vector, l: length of feature vector, CM: Confusion Matrix, ICA: Individual Class Accuracy, OCA: overall classification Accuracy,ICA_B : Individual class accuracy of Benign class, ICA_M: Individual class accuracy off Malignant class,B:Benign Class M: Malignant Class. The FV yielding the best OCA has been shaded with gray background.*

From Table 1, it can be observed that MFV yields higher OCA of 81.1 % in comparison to the SFV yielding the OCA of 79.2%. Further, it can also be observed that the MFV and SFV yield the same value for ICA benign class but for malignant class the ICA value of MFV is higher than ICA value obtained by SFV.

Table 2
Result of SVM classification of various individual Statistical features

FV	*l*	CM			OCA (%)	ICA_B(%)	ICA_M (%)
			B	M			
Edge	2	B	0	21	72.7	0.0	100
		M	0	56			
SFM	4	B	11	10	52.3	52.3	89.2
		M	6	50			
NGTDM	5	B	3	18	77.9	14.2	91
		M	3	53			
FOS	6	B	11	10	76.6	52.3	85.7
		M	8	48			
GLCM	13	B	12	9	79.2	57.1	87.5
		M	7	49			
GLRLM	11	B	10	11	77.9	47.6	89.2
		M	6	50			
GLDS	5	B	9	12	59.7	42.8	66.0
		M	19	37			

Note : ***CM :Confusion matrix , OCA : Over all classification accuracy , B: benign class , M: Malignant Class , ICA_B: Individual class accuracy of Benign class , ICA_M : Individual class accuracy of Malignant class, FOS : First order statistics , GLCM : Gray length co-occurrence matrix , GLRLM : Gray level run length matrix , GLDS: Gray level difference statistics SFM : Statistical feature matrix , NGTDM : Neighborhood gray tone difference matrix. The FV yielding the best OCA has been shaded with gray background.***

From table 2, it can be observed that the statistical feature vector consisting of GLCM features yield the highest OCA of 79.2 % and the ICA values of 57.1 % and 87.5 % for benign and malignant class respectively.

Table 3
Result of SVM classification of the combination of Morphological And various individual Statistical features

FV	*l*	CM			OCA (%)	ICA_B(%)	ICA_M (%)
			B	M			
MFV + Edge	11+2	B	14	7	84.4	66.6	91
		M	5	51			
MFV+ SFM	11+5	B	16	5	85.7	76.1	90.9
		M	6	50			
MFV+ NGTDM	11+5	B	13	8	83.1	61.9	91
		M	5	51			
MFV+ FOS	11+6	B	15	6	89.6	71.4	94.6
		M	5	51			
MFV+ GLCM	11+13	B	16	5	87.0	76.1	91
		M	5	51			
MFV+ GLRLM	11+11	B	15	6	87.0	71.4	92.8
		M	4	52			
MFV+ GLDS	11+5	B	13	8	88.3	61.9	98.2
		M	1	55			

Note ***: MFV: Morphological feature vector, l: length of feature vector, CM :Confusion matrix , OCA : Over all classification accuracy , B: benign class , M: Malignant Class , ICA_B: Individual class accuracy of Benign class , ICA_M : Individual class accuracy of Malignant class , MF: Morphological Features, FOS : First order Statistics , GLCM : Gray Length Co-occurrence Matrix , GLRLM : Gray level Run Length Matrix , GLDS: Gray level difference statistics SFM : Statistical Feature Matrix , NGTDM : Neighborhood Gray Tone Difference Matrix. The FV yielding the best OCA has been shaded with gray background.***

From table 3, it can be observed that the combined feature vector consisting of MFV + TFV consisting of FOS features yield the highest OCA of 89.6 % and the ICA values of 71.6 % and 94.6 % for benign and malignant class respectively

Table 4
Result of SVM classification of the combination of Morphological and all Statistical features

FV	*l*	CM			OCA(%)	ICA_B(%)	ICA_M(%)
			B	M			
MFV+SFV	11+46=55	B	14	7	83.1	66.6	89.2
		M	6	50			

Note: ***MFV: Morphological feature vector, SFV: Statistical feature vector l: length of feature vector, CM: Confusion Matrix , ICA : Individual Class Accuracy , OCA :overall classification accuracy , ICA_B : Individual class accuracy of Benign class , ICA_M : Individual class accuracy of Malignant class ,B: Benign Class M: Malignant Class***

From table 4, it can be observed that the combined feature vector consisting of MFV + SFV consisting of total 55 features yield the OCA value of 83.1 % and the ICA values of 66.6 % and 89.2 % for benign and malignant class respectively.

CONCLUSION

The exhaustive experiments carried out in the present work indicate it can be observed that the combined feature vector consisting of MFV + TFV consisting of FOS features yield the highest OCA of 89.6 % and the ICA values of 71.6 % and 94.6 % for benign and malignant class respectively. Thus it can be concluded that the shape based features along with texture features derived using first order statistics are significant to account for textural variations exhibited by sonographic appearance of benign and malignant breast. The proposed CAD system can assist practicing radiologists in decision making as second opinion tool and also reduce the need of unnecessary biopsies

ACKNOWLEDGEMENT

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

REFRENCES

1. Medical News TodayWhat is cancer? MNT Knowledge Center[online] [citied2015,October]. Availble from http://www.medicalnewstoday.com/articles/37136.php
2. Minavathi, Murali S , Dinesh M.S. " Classification of Mass in breast Ultrasound Images Using Image Processing Techniques" International Journal of Computer Applications ,2012; 42(10) : 29-36
3. Obajimi MO, Akute OO, Afolabi AO, Adenipekun AA, Oluwasola OO, Akang EEU, JoeRU , ATS Adeniji-Sofoluwe, Funmi O, Newstead G, Schmitt R andSennet C ,''BI-RADS Lexicon : An Urgent Call for the standardization of Breast Ultrasound in Nigeria" Annals of Ibadan Postgraduate Medicine, 2005 ; 3 (1): 82-88.
4. Alvarenga A.V. , Infantosi A. F. C. , Pereira W. C. A. and Azevedo C. M., "Assessing the performance of morphological parameters in distinguishing breast tumors on ultrasound images," Medical engineering & physics,2010; 32(1): 49–56.
5. Alvarenga A.V. ,Infantosi A.F.C., Pereira W. C.A. and Azevedo C.M.," Assessing the combined performance of texture and morphological parameters in distinguishing breast tumors in ultrasound images"2012 ;Med Phys 39(12):7350–7358.
6. Wu WJ, Moon WK " Ultrasound breast tumor image computer-aided diagnosis with texture and morphological features " 2008; AcadRadiol 15(7):873–880.
7. *Ultrasound Cases.info* , [online] ,2015[citied 2015,June] Available from http://www.ultrasoundcases.info/category.aspx?cat=67
8. *Image Processing and Analysis in JAVA* ,Image J 1.49 vesion1.6.024 *http://imagej.nih.gov/ij/download/win32/ij149- jre6-64.zip*
9. HuangY.-L., Chen D.-R. , JiangY.-R., Kuo S.-J. , Wu H.-K. , and Moon W. "Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound," Ultrasound in Obstetrics &Gynecology ;2008 , 32(4) : 565–572,.
10. Virmani J., Kumar V. , Kalra N. and Khandelwal N. "A rapid approach for prediction of liver cirrhosis based on first order statistics", in Proceedings of the IEEE International Conference on Multimedia, Signal Processing and Communication Technologies, IMPACT-2011, Aligarh, India, 2011 : 212-215.
11. Virmani J, Kumar V, Kalra N and Khandelwal N, "Prediction of cirrhosis based on singular value decomposition of gray level co-occurrence matrix and a neural network classifier", in *Proceedings of Development in E-systems Engineering, DeSE*, Dubai, 2011: 146-151

12. Vasantha M , SubbiahBharathi Vand Dhamodharan R. , “Medical image feature extraction, selection and classification”,International Journal of Engineering Science and Technology, 2010 ; 2, :2071-2076
13. Mohanaiah P , Sathyanarayanam PandGurukumar L, “Image texture feature extraction using GLCM approach”, *International Journal of Scientific and Research Publications*, 2013, 3: 1-5
14. Xu D.H.,Kurani A.S., Furst J.D. and Raicu D.S., “Run-length encoding for volumetric texture”, *Heart*, 2004 ; 27: 25-30.
15. Albregtsen F. , “Statistical texture measures computed from gray level run length matrices”, *Image*, 1995; 1: 3-8.
16. Weszka J.S., Dyer C.R. and Rosenfeld A. , “A comparative study of texture measures for terrain classification”, IEEE Transactions on Systems, Man and Cybernetics, 1976; 6: 269-285
17. Kim K., and Park H.W. , “Statistical textural features for detection of microcalcifications in digitized mammograms”, IEEE Transactions on MedicalImaging ,1999 ; 18.
18. Castellano G., Bonilha L., Li L.M. and Cendes F., “Texture analysis of medical images”, Clinical Radiology, 2004 ;59 :1061-1069
19. AmadasunM.andKing R. , “Textural features corresponding to textural properties”, IEEE Transactions on Systems, Man and Cybernetics,1989; 19 :1264-1274
20. Chang C.C., and Lin C.J., “LIBSVM, A library of support vector machines”.
21. Hassanien A.E.,Bendary N.E. , Kudelka M. and Snasel V., “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, 2011:269-279.
22. Azar A.T. and Said S.A. El., “Performance analysis of support vector machine classifiers in breast cancer mammography recognition”, Neural Computing and Applications,2014; 24:1163-1177
23. Jain S., Naik P. K., “System Modeling of cell survival and cell death : A deterministic model using Fuzzy System”, International Journal of Pharma and BioSciences(IJPBS),2012 ; 3(4): 358-373.
24. Virmani J., Kumar V.,Kalra N.andKhandelwal N., “SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors”, *Journal of Digital Imaging*, 2012 , 26(3) :530-543.
25. Virmani J.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, 2015; 381: 539-546.
26. Jain S., Chauhan D.S., “Mathematical Analysis of Receptors For Survival Proteins”, International Journal of Pharma and Bio Sciences (IJPBS), 2015, 6(3):164-176.