ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

# Classification of Breast Lesions Using the Difference of Statistical Features.

**Sahil Bhusri[*1], Shruti Jain[1], and Jitendra Virmani[2].**

[1]Jaypee University, Solan, Himachal Pradesh, India.
[2]Thapar University, Patiala, Punjab, India.

**ABSTRACT**

This paperclassifies thebreast lesions using the difference of statistical features obtained from outside area of interest (OAOI) and inside area of interest (IAOI).The breast lesions are differentiated into two classes benign and malignant. Texture featuresare computed using statistical texture feature modelsincluding SFM, NGTDM, FOS, GLCM, GLRLM, and GLDS.The SVM classifier has been used to classify the lesions on the basis of the features extracted from (*a*) OAOI, (*b*) IAOI and (*c*) the difference of statistical features computed from OAOI and the corresponding IAOI. The texture features computed using SFM texture feature model from OAOIsyields the maximum accuracy of 75% withindividual class accuracy values of 79.2% for benign and 72.2 % for malignant where the same texture features when computed from IAOIs yield the maximum classification accuracy of 65% with individual classification accuracyvalues of 45.8% and 77.7% for benign and malignant lesions respectively. However, the overall accuracy of 85 % is achieved using the difference of the GLRLM texture featuresbetween the OAOIs and IAOIs with individual classification accuracy values of 70.4 % for benign and 94.4 % for malignant lesions. Thus it can be concluded that the difference of GLRLM texture features computed from OAOIs and the corresponding IAOIs contain significant information for differential diagnosis between benign and malignant focal breast lesions.

**Keywords:** Breast cancer, Statistical features, Inside Area of Interest (IAOI), Outside Area of Interest (OAOI), Ultrasound.

***Corresponding author***

ISSN: 0975-8585

## INTRODUCTION

Human body is made up of million numbers of cells which group together to make a tissue or organ. Different types of tissues are present in different parts of body. Either these cells are replaced or reproduced at a regular interval of time but whenever uncontrolled division of cells is present, it leads to cancer. Breast cancer starts with the uncontrolled division of cells in the breast [1]. It is second most threatening disease in women that mostly proves to be fatal. It originates in the ducts or in the lobules of the breast. There is no existing method to prevent breast cancer but there are several ways that can help in the early detection of cancer. Various methods available are breast examination by physician (only at initial stage), X-ray, Ultrasonography and Magnetic Resonance Imagining (MRI). Biopsy is a surgical way in which a sample of tissues is taken from the abnormality for the analysis but it leads to the physical and physiological consequences on to the patient. To reduce the rate of biopsies, the most prescribed methods are Mammography and Ultrasonography. Mammography leads to high false report in case of young patients due to the dense tissue and has noisy types of images whereas the imaging by Ultrasonography offers non-radioactive, non-invasive, real time display and low cost as compared to the X-ray Mammography [2]. Various computer aided diagnosis systems have been introduced to make the diagnosis more explicit and distinct between benign and malignant tumors.

The abnormal change in the tissues due to any disease is called the lesion.There are various feature extraction and classification techniques available to classify the lesions of the breast. In this paper, analysis is done using the difference of the statistical features of *OAOI* and *IAOI*.

## METHODOLGY

The experimental flow for designing the proposed CAD system for classification between benign and malignant focal benign lesions in given in Figure 1.

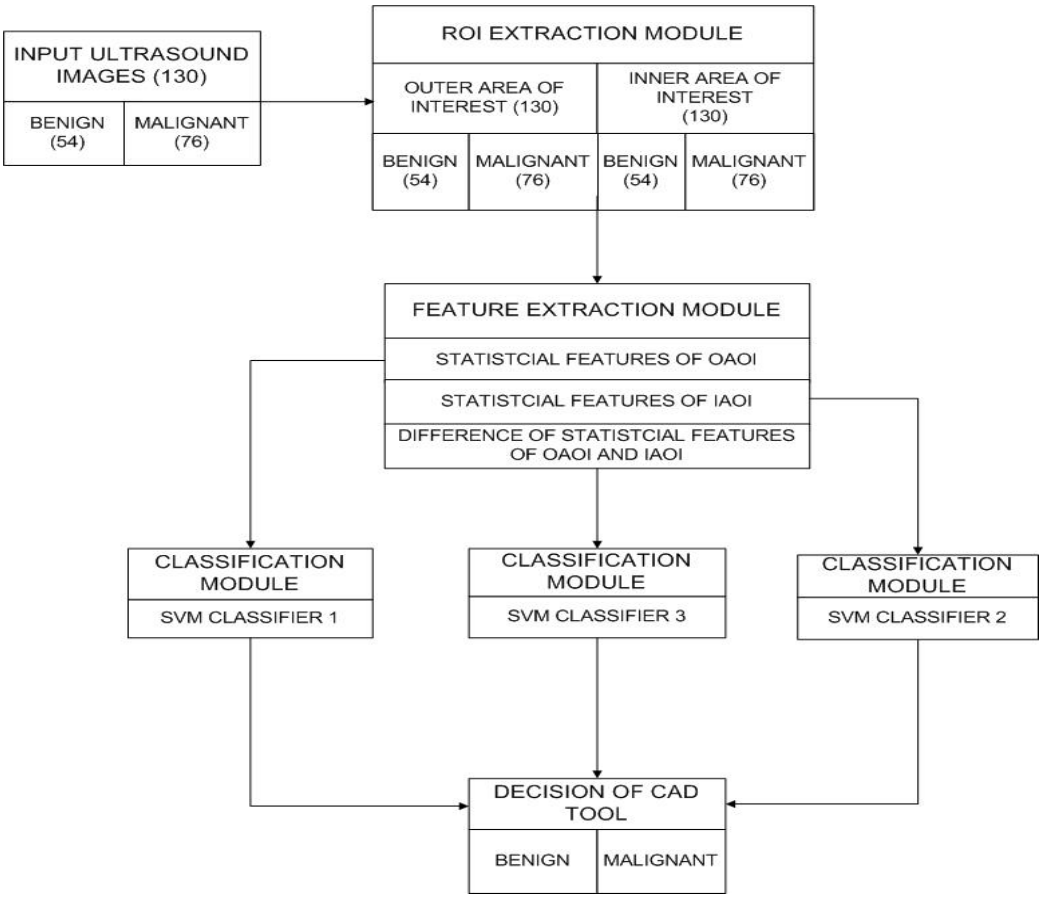


**Figure 1 : Experimental Work Flow: CAD system from benign and malignant focal breast lesions.**

ISSN: 0975-8585

***Ultrasound Image Database:***

The dataset of ultrasound images used for analysis is available online [3] and in this paper 130 cases having 54 cases of benign and 76 cases of malignant class are considered.

***Region of Interest (ROI) Extraction Module:***

The distortions in the breast ultrasound image are spotted under the directions of an experienced radiologist.The figure 2(*a*) and 2(*b*) indicate the samples images containing a benign and a malignant lesion. A variable size rectangular area of interest is taken from inside and outside the lesion. The *OAOI* contains the lesion and some surrounding tissue whereas *IAOI* contains the area inside the lesion .The figure 2(*c*) and 2(*d*) indicates the samples of outside and inside area of interest of benign case where as the figure 2(*e*) and 2(*f*) indicates the samples of outside and inside area of interest of malignant case.

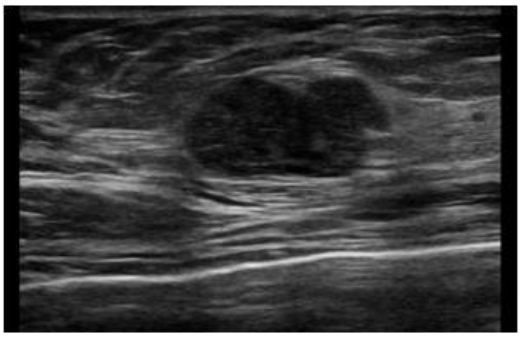

**Figure2(*a*): Sample of Benign Case**

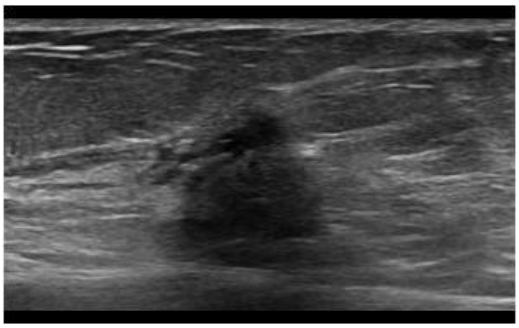

**Figure 2(*b*) : Sample of Malignant case**

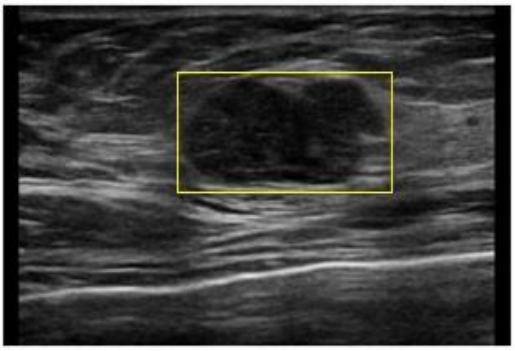

**Figure 2(*c*): Outside Area of Interest (OAOI) of Benign case.**

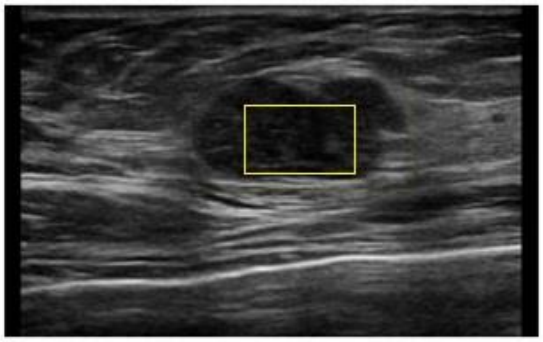

**Figure 2(*d*): Inside Area of Interest (IAOI) of Benign case.**

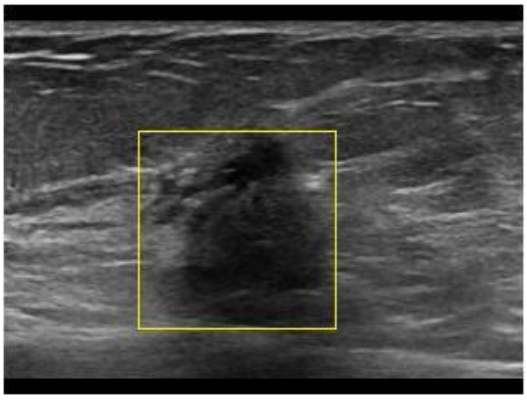

**Figure 2 (*e*): Outside Area of Interest (OAOI) of Malignant Lesion.**

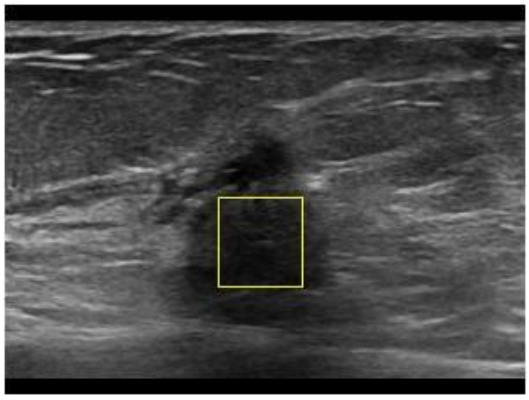

**Figure 2*(f)*: Inside Area of Interest (IAOI) of Malignant Lesion.**

ISSN: 0975-8585

***Feature extraction module***

The lesions of breast can be characterized by computing various mathematical features i.e. *morphological features and texture features*. Morphological features extract the shape based features whereas the texture features helps in characterization by defining the properties of the surface. These features are computed by three methods i.e., (*a*) Signal processing based methods (*b*) Transform domain Methods (*c*) Statistical Methods. In this paper statistical feature models have been used to compute the texture features from OAOI and IAOI. Difference of the statistical features is the difference between statistical features of OAOI and IAOI.As an ultrasound image contains pixels of different gray level intensities and based on the divisions of these gray level intensities, statistical features are computed using first order statistics, second order statistics and higher order statistics i.e. SFM, NGTDM, FOS, GLCM, GLRLM, and GLDS are computed [4-6]. Different statistical features with different feature length are shown in figure3.

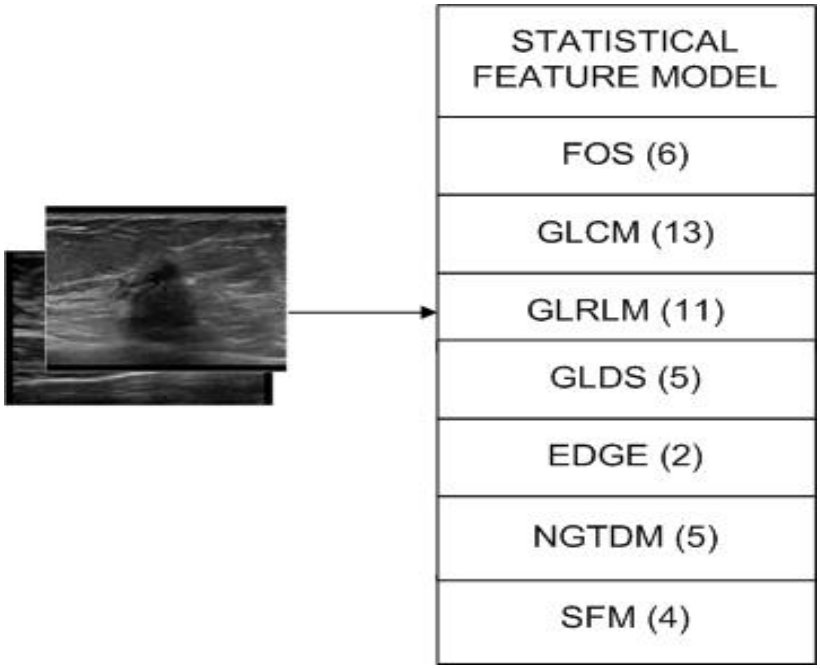


**Figure 3: Different Statistical Feature Models used in this work**

First order statistics: *This technique uses image histogram's moments to describe the texture. They compute the randomness, roughness, uniformity and entropy and average gray level [7].*

*Second order statistics:* The computations with the Gray level co-occurrence matrix (GLCM) were included in the second order statistics. GLCM gives the combinations of pairs of pixels that occur frequently having different gray level occurring in an image. They have separation of different dimensions indifferent directions of θwhere it can take values of 0°, 45°, 90°, 135°. A total of 13 GLCM features are computed[8-10].

*Higher order statistics:*Higher order statistics are computed with the use of Gray Level Run Length Matrix (GLRLM).Texture features are computed taking the different combinations of intensities present at relative position from one another. The set of consecutive pixels of same gray levels defines the gray level run andno of times a run occurs tells about the run length. A total of 11 GLRLM features are computed [11-12]. Here *i* denotes the gray levels, *j* denotes the run length and$P(i, j|\vartheta)$ denotes the no of occurrences of run length *j* at gray level*i* in the direction of $\vartheta$.

*Other Statistical Features*: The Gray Level Difference Statistics (GLDS) [13-14], Edge, SFM and NGTDM features are the other statistical features are computed for the present classification task.GLDS calculatescontrast, energy, entropy, homogeneity, and meanon the basis of the co-occurrence of a pixel pairs that have difference in gray levels. Edge feature computes information present in edges image Edge features computes Absolute gradient mean and absolute gradient variance The gradient will be high if there are abrupt changes present

else it will be low. SFM (Statistical Feature Matrix) calculates coarseness, contrast, periodicity and roughness of pixels at distinct distances within an image [15]whereas NGTDM considers the difference between the gray level of the pixels [16] and computes busyness, coarseness, complexity, contrast, and strength.

***Classification Module:***

The classification is the process of providing label to images to predefined corresponding classes. The dataset is already defined in case of supervised learning and is not available in case of unsupervised learning.Support Vector Machine (SVM) classifier is a supervised learning machine and works on basis of statistical theory. SVM classifier supports both linear and non - linear classification. Support vector machines creates ahyper plane between the classes with the help of the training data available and good separation is achieved intuitively but the sets that are available to discriminate are not linearly separable in the space. Therefore in non - linear classification problems, the input data is mapped in to the kernel functions in which the data is mapped from input space to the higher dimensional feature space.For the classification task, Gaussian radial basis kernel's function has been used. Recent algorithms include the sub gradient and coordinate descent methods that have a big plus of having large and sparse datasets. LibSVM library is included for the implementation of SVM classifier [17-31].

## RESULTS AND DISCUSSION

In this paper a total of 130 cases were considered out of which 54 cases are of benign class and 76 cases are of malignant class. A variable size ROI was selected inside and outside the lesion.Statistical feature model is used to extract features from both of the areas of interest. Difference of the statistical features is taken as the difference between statistical features of OAOI and IAOI. The list of different experiments carried out in this paper regarding OAOI, IAOI and the difference of statistical features is tabulated in given Table 1 whereas the results of these experiments are tabulated in given Table 2 to Table 4.

**Table1: List of Experiments carried out to classify the Breast lesions**

| | |
|---|---|
| EXP 1 | To obtain the classification performance of texture features obtained from IAOI |
| EXP 2 | To obtain the classification performance of texture features obtained from OAOI |
| EXP 3 | To obtain the classification performance of texture features obtained using the difference of statistical features of OAOI and IAOI |

The classification performance of different FVs obtained from IAOIs using SVM classifier is given in Table 2.

**Table 2 Classification performance of different FVs obtained from IAOIs using SVM classifier**

| FVs | *l* | CM | | | OCA (%) | $ICA_B$(%) | $ICA_M$ (%) |
|---|---|---|---|---|---|---|---|
| Edge | 2 | | B | M | 45 | 37.5 | 50 |
| | | B | 9 | 15 | | | |
| | | M | 18 | 18 | | | |
| SFM | 4 | B | 11 | 13 | 65 | 45.8 | 77.7 |
| | | M | 8 | 28 | | | |
| NGTDM | 5 | B | 8 | 16 | 65 | 33.3 | 86.1 |
| | | M | 5 | 31 | | | |
| FOS | 6 | B | 7 | 17 | 60 | 29.1 | 80.5 |
| | | M | 7 | 29 | | | |
| GLCM | 13 | B | 12 | 12 | 48.3 | 50 | 47.2 |
| | | M | 19 | 17 | | | |
| GLRLM | 11 | B | 3 | 21 | 56.6 | 12.5 | 86.1 |
| | | M | 5 | 31 | | | |
| GLDS | 5 | B | 13 | 11 | 55 | 54.1 | 88.1 |
| | | M | 4 | 32 | | | |

**Note :** FVs: Feature Vectors,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 having the best OCA has been shaded with gray background.*

From Table 2 it can be observed that SFM feature has the highest OCA of 65 % with the ICA values of 45.8 % for benign and 88.1 % for malignant class, obtained using SFM feature.

The classification performance of outside area of interest (OAOI) obtained by various individual feature vectors is depicted in Table3.

**Table 3 : Classification performance of different FVs obtained from OAOIs using SVM classifier**

| FVs | *l* | | CM | | OCA (%) | $ICA_B$ (%) | $ICA_M$ (%) |
|---|---|---|---|---|---|---|---|
| Edge | 2 | | B | M | 66.6 | 50 | 77.7 |
| | | B | 12 | 12 | | | |
| | | M | 8 | 28 | | | |
| SFM | 4 | B | 19 | 5 | 75 | 79.1 | 72.2 |
| | | M | 10 | 26 | | | |
| NGTDM | 5 | B | 3 | 21 | 58.3 | 12.5 | 88.8 |
| | | M | 4 | 32 | | | |
| FOS | 6 | B | 10 | 14 | 46.6 | 58.3 | 50 |
| | | M | 18 | 18 | | | |
| GLCM | 13 | B | 8 | 16 | 70 | 33.3 | 94.4 |
| | | M | 2 | 34 | | | |
| GLRLM | 11 | B | 13 | 11 | 71.6 | 54.1 | 88.8 |
| | | M | 4 | 32 | | | |
| GLDS | 5 | B | 10 | 14 | 51.6 | 41.6 | 58.3 |
| | | M | 15 | 21 | | | |

**Note :** FVs: Feature Vectors,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 having the best OCA has been shaded with gray background.*

From Table 3 it can be observed, that SFM feature has the highest OCA of 75 % with the ICA values obtained for benign and malignant classes as 79.1 % and 72.2 % respectively,with SFM feature.

The classification performance of difference of statistical features outside area of interest (OAOI) and inside area of interest (IAOI) is depicted in Table 3.

**Table 4 : Classification performance of difference of FVs obtained from of OAOI and IAOI using SVM classifier.**

| FVs | *l* | CM | | | OCA (%) | $ICA_B$(%) | $ICA_M$ (%) |
|---|---|---|---|---|---|---|---|
| Edge | 2 | | B | M | 60 | 79.1 | 47.2 |
| | | B | 19 | 5 | | | |
| | | M | 19 | 17 | | | |
| SFM | 4 | B | 13 | 17 | 45 | 54.1 | 38.8 |
| | | M | 22 | 14 | | | |
| NGTDM | 5 | B | 4 | 20 | 61.6 | 16.6 | 91.6 |
| | | M | 3 | 33 | | | |
| FOS | 6 | B | 7 | 17 | 53.3 | 29.1 | 69.4 |
| | | M | 11 | 25 | | | |
| GLCM | 13 | B | 9 | 15 | 66.6 | 37.5 | 86.1 |
| | | M | 5 | 31 | | | |
| GLRLM | 11 | B | 17 | 7 | 85 | 70.8 | 94.4 |

ISSN: 0975-8585

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | M | 2 | 34 | | | |
| GLDS | 5 | B | 11 | 13 | 59.7 | 45.8 | 80.5 |
| | | M | 7 | 29 | | | |

**Note :** FVs: Feature Vectors, 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 having the best OCA has been shaded with gray background.*

From Table 4 it can be observed that GLRLM feature has the highest OCA of 85 % and the ICA values obtained for benign and malignant classes are 70.8 % and 94.4 % respectively.

The figure 4 represents the proposed CAD system design for differential diagnosis between breast lesions.

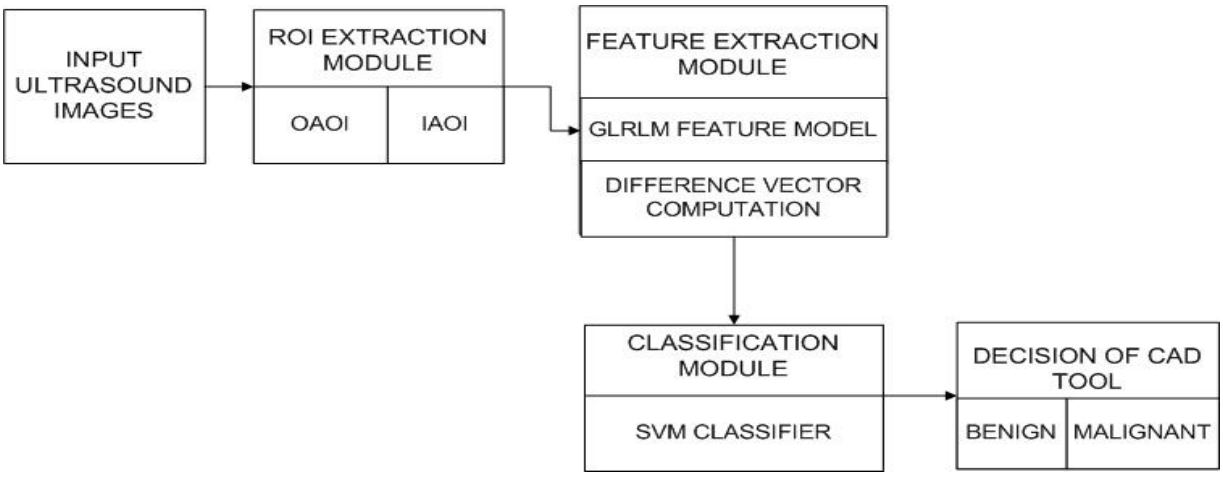


**Figure 4: Proposed CAD System**

Figure 4 represents the proposed CAD system for the classification of breast lesions. It shows that after extracting the OAOI and IAOI, the only need is to use GLRLM feature model to extract features and take the difference of the GLRLM features which will not only reduce the overheads but also save time . The difference of GLRLM feature is further passed on to SVM classifier for classification.

## CONCLUSION

This paper proposes a CAD system for the radiologists as a second opinion tool for the breast cancer using GLRL texture feature difference vector. The experiments carried out in the present work signify that the GLRL texture difference vector computed from OAOI and corresponding IAOI yields the maximum OCA of 85%for differential diagnosis between benign and malignant breast lesions using the ultrasound images with individual class accuracy values of 70.8 % and 90.4 % for benign and malignant lesions respectively.

## 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] Cancer Research UK[online] ,2016 [cited 2016 ,March]http://www.cancerresearchuk.org/about-cancer/what-is-cancer/how-cancer-starts

[2] Minavathi Murali S, Dinesh MS , International Journal of Computer Applications ,2012; 42: 29-36.

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

[4] Bhusri S, Jain S and Virmani J, in Computing for Sustainable Global Development (INDIACom), 3rd International Conferences, March 2016 : 2523-2527.

[5] Rana S , Jain S and Virmani J , in Computing for Sustainable Global Development (INDIACom), 3rd International Conferences, March 2016 : 2528-2532.

[6] Bhusri S , Jain S and Virmani J, International Journal of Pharma and BioSciences, 2016 April; 7:(B) 617 – 624.

[7] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, in Proceedings of the IEEE International Conference on Multimedia, Signal Processing and Communication Technologies, IMPACT-2011, Aligarh, India, 2011 : 212-215.

[8] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, in Proceedings of Development in E-systems Engineering, DeSE, Dubai, 2011: 146-151.

[9] M. Vasantha, V. Subbiah Bharathi and R. Dhamodharan, International Journal of Engineering Science and Technology, 2010 ; 2 :2071-2076

[10] P. Mohanaiah, P. Sathyanarayanam and L. Guru Kumar, 2013, 3: 1-5

[11] D.H. Xu, A.S. Kurani, J.D. Furst and D.S. Raicu, “, Heart, 2004; 27: 25-30.

[12] F. Albregtsen., Image, 1995;1: 3-8.

[13] J.S. Weszka, C.R. Dyer and A. Rosenfeld, IEEE Transactions on Systems, Man and Cybernetics, 1976; 6: 269-285.

[14] J.K. Kim and H.W. Park, IEEE Transactions on MedicalImaging, 1999; 18.

[15] G. Castellano,L.Bonilha, L.M. Li and F. Cendes, Clinical Radiology, 2004 ; 59 :1061-1069

[16] M. Amadasun and R King , IEEE Transactions on Systems, Man and Cybernetics, 1989; 19 :1264-1274.

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

[18] S Jain, DS Chauhan, International Conference on Information and Communication Technology for Sustainable Development (ICT4SD), Ahmedabad, India, July 3-4 2015 ; 81-88.

[19] A. E. Hassanien , N.E. Bendary,M. Kudelka and V. Snasel , in Proceedings of 3rd International Conference on Intelligent Human Computer Interaction IHCI 2011, Prague, Czech Republic, 2011:269-279.

[20] S Jain, Control System and Power Electronics – CSPE 2015, Bangalore., Aug 1-2, 2015

[21] S Jain, PK Naik, SV Bhooshan , International Conference on Computational Intelligence and Communication Networks (CICN2011), Gwalior, India. Oct 07-09, 2011; 565-568.

[22] S. Jain,P.K. Naik, International Journal of Pharma and BioSciences(IJPBS), 2012;3: 358-373.

[23] J. Virmani,V. Kumar, N. Kalraand N. Khandelwal, Journal of Digital Imaging, 2012 , 26 :530-543.

[24] S.Jain, D.S. Chauhan, International Journal of Pharma and Bio Sciences (IJPBS), 2015, 6:164-176.

[25] J. Virmani and Kriti , in Proceedings of 2nd International Conference on Computer and Communication Technologies (IC3T 2015),24-26 July, 2015, Hyderabad, 2015, 381: 539-546.

[26] S Jain,Research Journal of Pharmaceutical, Biological and Chemical Sciences (RJPBCS),2016 ; 7

[27] J. Virmani, V. Kumar, N. Kalra, N. Khandelwal, Journal of Digital Imaging, (2013) 26(6): 1058-1070

[28] Jain S, Network Biology, June 2016; 6:40-46

[29] J. Virmani, V. Kumar, N. Kalra, N. Khandelwal, Defence Science Journal, 63

[30] Kriti, J. Virmani, N. Dey, V. Kumar, Applications of Intelligent Optimization in Biology and Medicine, Intelligent Systems Reference Library 96, (2016) : 159-180.

[31] Jain S, Research Journal of Pharmaceutical, Biological and Chemical Sciences, May- June 2016; 7