I J C T A, 9(17) 2016, pp. 8811-8815

Malignant and Benign Tumor using SVM Classifier

Navjot Kaur*, Sanjay Singla* and Ashima Goyal*

ABSTRACT

Breast cancer, the second common cancer after lung cancer realized in females if compared with males.GLCM features are energy, correlation, contrast and homogeneity. In this proposed work, features which realized by GLCM called Gray Level Co-Occurrence Matrix and comparison between the features having original image values and their corresponding images values having cancer are done.DSM(Digital Database for screening Mammography) and MIAS(The Mammographic Image Analysis Society) images are used to expose the benign and malignant cancer by various digital image processing techniques. For classification of Malignant and Benign cancer, SVM Classifier has been used as having good accuracy.

Index Terms: Malignant, Benign, DDSM/MIAS, GLCM Features, SVM Classifier.

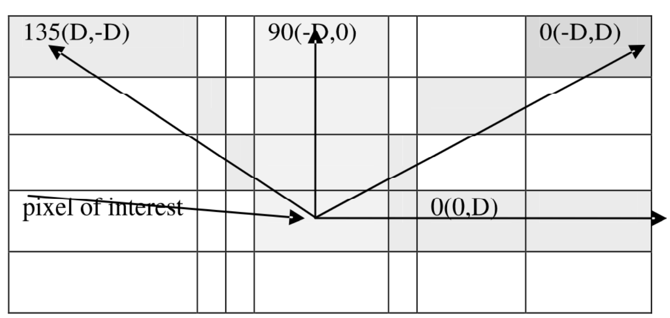
I. INTRODUCTION

Malignant and Benign cancer are the most common cancer after the lung cancer. These cancers are the types of breast cancer.GLCM, 2nd order statistics, is used to collect the information about pixels pairs and also used to allow the pixel brightness of an image.Haralick put forward for considerations different features. These features are calculated at separate angles. These angles are at [1,-1], [-1,0], [0,1], [-1,1]. GLCM order texture measures are First order texture features, Second order texture ,Third and higher order textures features. [2]GLCM is calculated from gray scale image values only. It cannot be applied on RGB images.

GLCM CALCULATION: GLCM is calculated with the help of sub –region image.

GLCM

1	2	0	0	1	0	0	0
0	0	1	0	1	0	0	0
0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	0
1	0	0	0	0	1	2	0
0	0	0	0	0	0	0	1
2	0	0	0	0	0		0
0	0	0	0	1	0	0	0



SUB-REGION/IMAGE

1	1	5	6	8
2	3	5	7	1
4	5	1	1	2
8	5	1	2	5

* Department of Computer Science and Engineering, IET BHADDAL, Ropar, Punjab, India, *E-mail: navjotdhadli76@gmail.com; hodcse@ietbhaddal.edu.in; ashima.goyal@ietbhaddal.edu.in*

(A) GLCM Features

There are various types of GLCM features Energy, Correlation, Contrast and Homogeneity are the GLCM features which are extracted. With the help of weight of the equation, the texture measures are used and according to the degree, the texture is classified. The term having square leads the second order equation and the term having cube shows the third order equation [6,7]The features which are extracted are:

1. Contrast: Sum of Square Variance is called Contrast and contrast features calculates the intensity and the Range is [0, size (GLCM, 1)-1)^2] [2], [3], [5], [6]

$$\sum_{i,j=0}^{N-1} Pi, j\,(i-j)\wedge 2$$

For horizontal GLCM,

Contrast =contrast weight is multiplied by horizontal GLCM.

1. This feature calculates the pixel correlation and its neighbor over the complete image. For the Constant image, its value is NaN. and range Range=[-1,1] and the formula is

$$\sum_{i,j=0}^{N-1} Pi, j\,\frac{(i-\mu i)(j-\mu j)}{\sqrt{(\sigma i^2)(\sigma j \wedge 2)}}$$

2. Energy: The addition of elements having in GLCM is Energy. The square root of ASM(Angular Second Moment is called Energy and the range of it is[0 1].The constant image has its value 1.[2],[6],[7] and the equation is

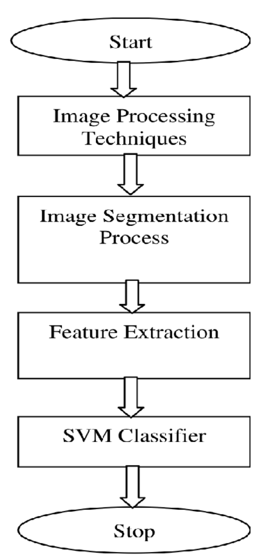
$$\sum_{I,J=0} P(i,j)\wedge 2$$

3. Homogeneity: The closeness how the elements in GLCM distribute in the GLCM to the GLCM having diagonal is calculated by Homogeneity. GLCM value is 1 and its range is [0,1] for diagonal elements. The equation which shows it is :

$$\sum_{i,j=0}^{N-1} P(i,j)/R$$

Here R=1/1+ (i-j)^2.,[6],[7].Comparison between the original image feature values and its related images having cancer have been done and the result is there shown in Table 1.

FLOW CHART



(B) SVM Classifier

SVM is called the Support Vector Machine and this classifier is used to classify the images which depicts that which portion of the image is malignant or benign. It is effective for higher dimensional images. The cases when samples are less than the number of dimensions this classifier is also operative. Its accuracy is good than KNN classifier.SVM Classifier is used to classify the all pixels of the image. This depicts that this division of the image are of two regions and the boundary between these two regions is called the optimal separating hyperplane. This hyperplane is used to differentiate between two classes whether the cancer is malignant or benign. In this Support Vector classification, the function which is used is a combination of kernels linearly associated with the Support Vectors .This formulation has advantage that the SVM training is required for solving a problem having quadratic optimization which is used for optimization routines from libraries having numerical values. This step is fast computationally and problem regarding stability is there [5]. Other Attractive iterative algorithms i.e. the Sequential Minimal Optimization (SMO), Nearest Point Algorithm (NPA) etc. are there to defeat this problem [4], [5]. SVM Classifier has accuracy of 95% approximately.

II. PROCEDURE

1. The original image is taken.The image can be Digital database for screening mammography or The Mammographic Image Analysis Society .
2. The DDSM/MIAS image is changed into the Gray Scale as GLCM Features extract on gray scale images only.
3. Cropping method is used to remove the noise like artifacts. It is optional as no artifact is present.
4. Filtering is done and filter used here is median filter as filtration is good than others.
5. Thresholding is the another process which is used on filtered image.
6. Segmenttaion is the another process which is used after thresholding of the image that segments the threshold image.[4],[5]
7. Classifier is the main requirement to classify whether the cancer is malignant or benign.SVM Classifier is used to classify the image as its accuracy is good than KNN classifier etc.

   Here PN is used for Patient. The results of the various images are:

(A) IMAGE DDSM (PNT1)

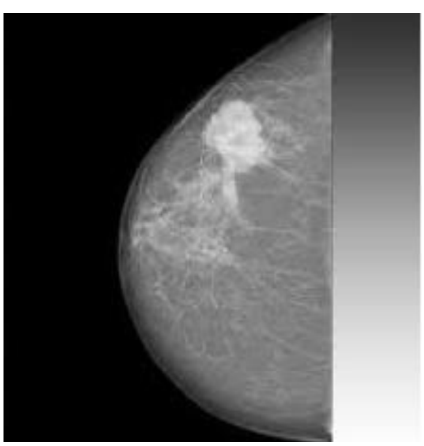


Figure 1(aa): Original Image

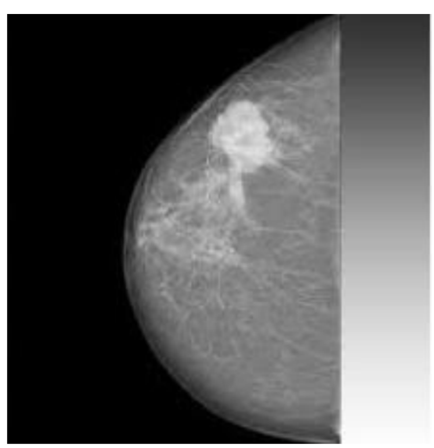


Figure 1(ab): Cropped image

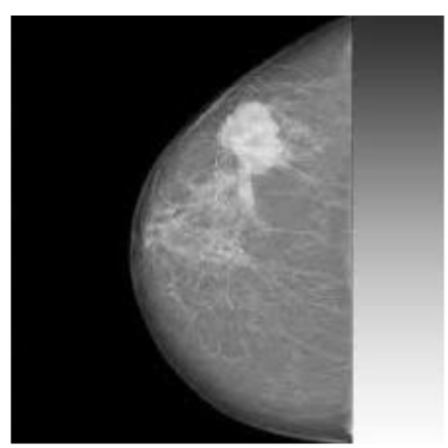


Figure 1(ac): Gray Scale Image

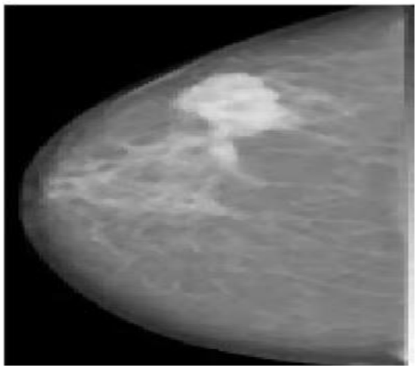


Figure 1(ad): Filtered Image

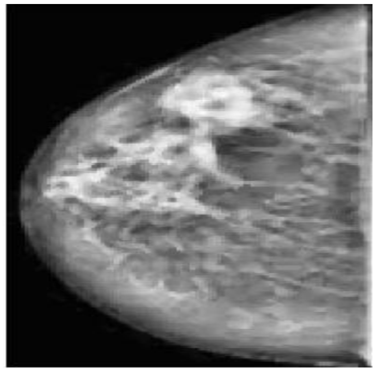


Figure 1(ae): Contrast Image

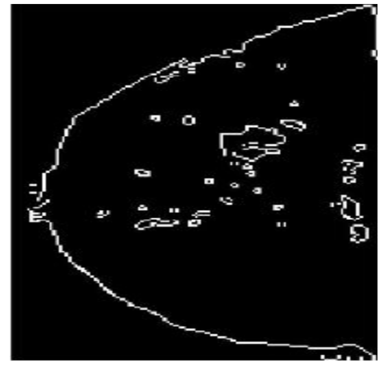


Figure 1(af): Segmented Image

(B) IMAGE DDSM (PNT2)

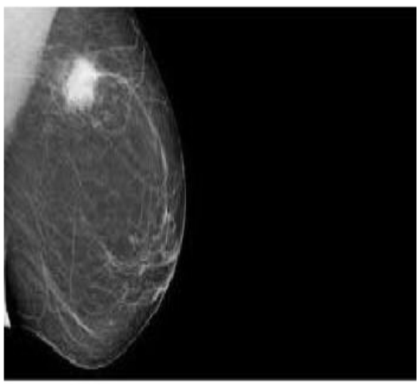


Figure 2(aa): Original Image

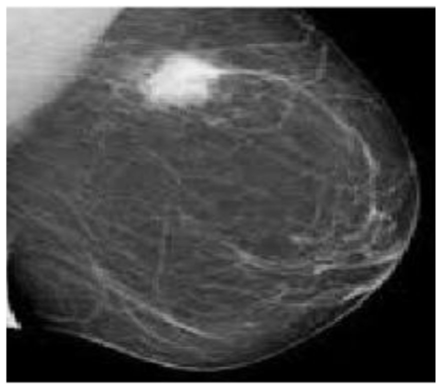


Figure 2(ab): Cropped Image

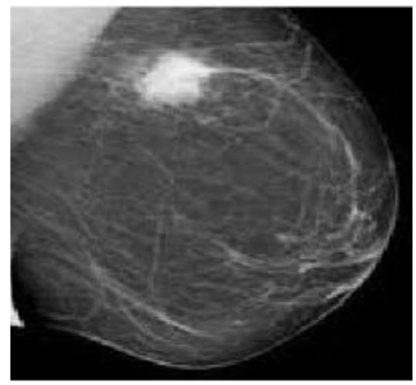


Figure 2(ac): Gray Scale Image

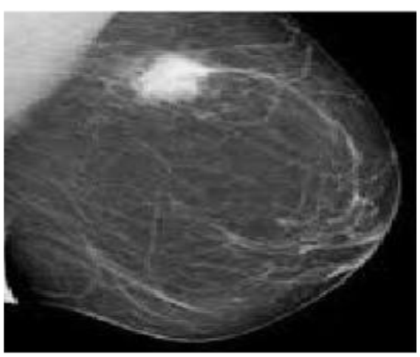


Figure 2(ad): Median Filterd Image

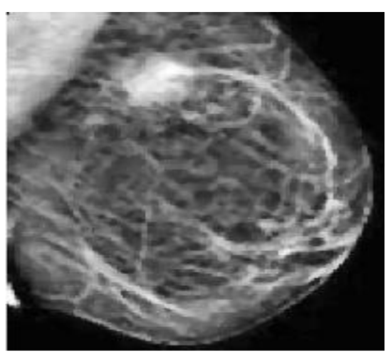


Figure 2(ae): Contrast Image

Figure 2(af): Segmented Image

III. CONCLUSION AND RESULT

To detect the breast cancer, database images whether DDSM or MIAS are taken and apply some MATLAB Version R200b programming using digital image processing. It includes pre-processing technique for noise reduction and image enhancement. Image segmentation used to detect the segmented part and then extract the features of GLCM.

By extracting GLCM features for Mammogram, we concluded that except correlation, the other features Contrast, Energy and Homogeneity all have larger values for Original Images as compared to their Cancerous images.

Table 1
GLCM Texture Features Values

Texture Features	*PNT1 OV*	*PNT1 CV*	*PNT2OV*	*PNT2CV*
Offset[0 1]				
Contrast	0.4212	0.4869	.1200	.7626
Correlation	0.9594	0.842	.9685	.9588
Energy	0.2475	0.9275	.4099	.4865
Homogeneity	0.9523	0.9923	.9756	.9864
Offset[-1 1]				
Contrast	0.9831	0.9866	.1832	1.0501
Correlation	0.7727	0.7014	.9519	.9570
Energy	0.2319	0.9176	.4056	.4809
Homogeneity	0.8794	0.9843	.9665	.9812
Offset[-1 0]				
Contrast	0.6646	0.4983	.1011	.7403
Correlation	0.9195	0.8881	.9609	.9697
Energy	0.2343	0.3965	.4091	.4768
Homogeneity	0.8836	0.9921	.9771	.9878
Offset[-1-1]				
Contrast	0.9909	1.1906	.1794	.9596
Correlation	0.8616	0.7575	.9529	.9604
Energy	0.2216	0.8925	.4059	.4823
Homogeneity	0.8715	0.9804	.9680	.9837

OV-Original Image Value

CV-Cancerous Image Value

PNT-Patient

If the size of images increases, all the feature values increase whether it is the case of original images or cancerous images. Image size can be of 64x64, 128x128, 256x256 etc. Out of these the optimal value of image is taken i.e.128x128.

REFERENCES

[1] D. NarainPonraj, M.K. Evangelin Jennfier, P Poongodi, J Samuel Manoharan, "A Survey on the Preprocessing Techniques of Mammogram for the Detection of Breast Cancer*" Journal of Eme rging Trends in Computing and Information Sciences,* vol. 2, NO. 12, December 2011.

[2] R. Ramani, Dr. N. Suthanthira Vanitha, S. Valarmathy," The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images", *I.J. Image, Graphics and Signal Processing,*vol 5,pp. 47-54,2013.

[3] Blaine Martinez, Haidi Ibrahim, Nicholas SiaPik Kong, Theam Foo Ng,"Simple Adaptive Median Filter for the Removal of Impulse Noise from Highly Corrupted Images*",IEEE Transactions on Consumer Electronics*, Vol 54, pp .4, Novermber 2008.

[4] Blaine Martinez," Digital Image Processing Final Project" and Nicholas Sia Pik Kong, Theam Foo Ng ,"Simple Adaptive Median Filter for the Removal of Impulse Noise from Highly Corrupted Images HaidiIbrahim*", IEEE Transactions on Consumer Electronics*, Vol 54, No.4 ,Novermber 2008

[5] Robert M. Haralick ,K. Shanmugam and Its'hak Dinstein,"Textural Features for Image Classification", *IEEE Transcations on systems,man and cybernetics,v*ol.SMC-3 pp. 610-621,6-Nov 1973.

[6] Pawar B.V., Patil Sharma,"Early Stage Detection ofTumors in Mammograms",Electronic Dept.,W.C.E. Sangli.

[7] Aswini Kumar Mohanty, Swapnastikta Beberta,S aroj Kumar Lenka," Classifying Benign and Malignant Mass using GLCM and GLRLM based Features from Mammogram", *International Journal of Engineering Research and Applications*,col.1,pp.6876-693

[8] P. Mohanaiah, P. Sathyanarayana, L. Guru Kumar," Image Texture Feature Extraction Using GLCM Approach", *International Journal of Scientific and Research Publications,* Volume 3, Issue 5, May 2013.