ISSN: 0976-3104
SPECIAL ISSUE (SCMDSA)
Kamalakannan and Babu

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

EARLY DETECTION OF BREAST CANCER USING GLCM FEATURE EXTRACTION IN MAMMOGRAMS

Kamalakannan J[1]* and Rajasekhara Babu. M[2]

[1]*School of Information Technology and Engineering, VIT University Vellore, Tamil Nadu, INDIA*

[2] *School of Computing Science and Engineering, VIT University, Tamil Nadu, INDIA*

ABSTRACT

Breast cancer is the one of the most common invasive cancer type among women. Early detection and diagnosis of breast cancer can be facilitating the chance of better treatment for the cancer affected people with mammography image analysis, since mammograms are cost effective and the world standard for screening of breast. Extracting the features from mammograms will help in identifying and classifying the breast abnormalities. There are many ways to extract the features; in this paper we have used GLCM to extract features from the mammographic images. GLCM is a statistical method of examining texture that uses the spatial relationship of pixels [6] . the features which are extracted can be given to a classifier to classify the abnormalities as benign and malignant. The mammograms from mini MIAS database is used for extracting the features. The radiologist uses the CAD system for differentiating benign and malignant abnormalities from the mammograms in a better way. The technique which is adapted in this paper can be helpful in improving the performance of the CAD system which can assist the radiologist for better diagnosis of breast cancer.

Received on: 30th-Nov-2015
Revised on: 25th-March-2016
Accepted on: 29th – March-2016
Published on: 23rd–May-2016

KEY WORDS

Mammogram; GLCM; Benign; Malignant; CAD; Screening; Feature

***Corresponding author: Email:** jkamalakannan@vit.ac.in **Tel:** +91-9944730423

INTRODUCTION

Today, Breast cancer is one of the common among cancer both men and women. Breast cancer is the most common invasive cancer in females worldwide. It accounts for 16% of all female cancers and 22.9% of invasive cancers in women.18.2% of all cancer deaths worldwide, including both males and females, are from breast cancer. According to the National Cancer Institute, 232,340 female breast cancers and 2240 male breast cancers are reported in the USA each year and 39,620 caused by the disease as well [1]. The breast cancer is the most affecting cancer in women compared to other types of cancer. The risks of the breast cancer increases with the factors such as female gender, obesity, lack of physical exercise , having children late or not at all etc .It has been found that the 80% of women are above age of 50[1]. Breast cancer can be easily diagnosed with various techniques. Imaging tests use x-rays, magnetic fields, sound waves, or radioactive substances to create pictures of the inside of your body. The process of examination of breast to identify the abnormality is called mammography. It is recommended that women of age 40 and older have regular mammogram to detect the breast cancer at early stage. The gold standard and cost effective way of screening the breast cancer is through mammograms [2].

Mammogram

A mammogram is one of the best radiographic methods to detect the breast cancer at early stage. It detects the tumors which are tiny and it is very difficult to identify by the radiologist . Mammography gives us the X-ray image as an output [3]. Image Processing techniques that provides a sufficient assessment to category the abnormalities[3] such as calcification(a),circumscribes masses (b),speculate masses(c),ill-defines masses (d),Architectural distortion(e), asymmetry (f) to make a clear diagnosis of the images[3]. The Current usage of early detection of breast cancer is done through mammography screening [4]. Mammogram is a medicinal practice for distinguish the breast growth which was initially coined by Bob Eagan in 1950. Mammogram is the radiology tool which gives better accuracy than clinical breast examination [4]. It not only identifies the abnormalities but also identify the normal breast among women [4]. This Detection strategy is termed as

ISSN: 0976-3104

mammography, in which X-beams of low vitality will be anticipated on an emulsion film that gives a white washed duplicate which symbolizes the tissue in the bosom [4]. Basically, there are two sorts of perspectives in a mammography namely crania-caudal view (CC) and Mediolateral Oblique (MLO).Earlier view is normally recognized in both diagnostics test using mammogram and the clinical breast examination [4]. In this viewpoint, we can see maximum conceivable vicinity of a granular tissue, the adjoining greasy tissue and edge of the midsection divider muscle [4]. Later view is considered for the routine mammogram. Cumbersome region is additionally given by CC view than by MLO view which are shown in **Figure- 1**.

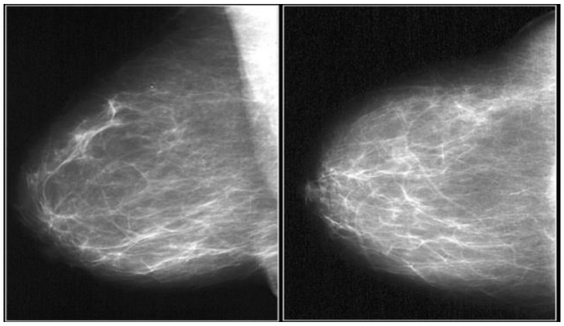

Fig: 1. MLO and CC views of the same breast

When we narrow down the perspective of mammogram, Later medial perspective are viewed from the outside towards the focal point whereas mediolateral perspective are viewed from the inside portion of breast [4]. There are different kinds of tumor may present in the mammogram. The tumor with speculated shape will be the cancerous tumor (Malignant) and the tumor with circular shape will be the noncancerous tumor (Benign). The masses with different shape and margin are depicted in the **Figure- 2**.

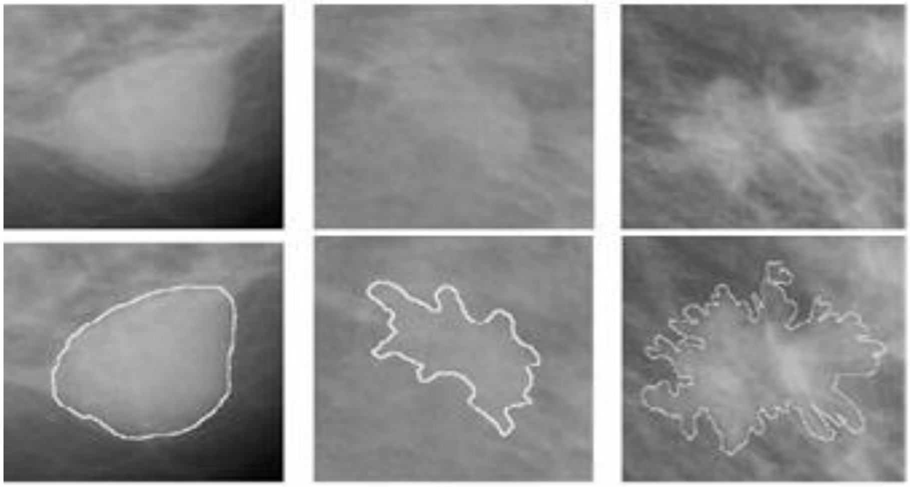

Fig: 2. Three mass examples with different shape and margin: (a) circular shape and circumscribed margin, (b) lobular shape and well defined margin, and (c) speculated shape and ill-defined margin. The last of the three has a higher malignancy probability.

The Grey Level Co-occurrence Matrix (GLCM) Features

Grey-Level Co-occurrence Matrix (GLCM) texture measurements is the one of the way to extract features for image texture since they were proposed by Haralick [5], and 14 statistical features were introduced. GLCM is a statistical method of examining texture that uses the spatial relationship of pixels [6]. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image. These features are generated by calculating the features for each one of the co-occurrence matrices obtained by using the directions 0°, 45°, 90°, and 135°[7] , then averaging these four

THE IIOAB JOURNAL
ISSN: 0976-3104

values .The GLCM is a intensity change histogram as a function of distance and direction. It is an estimate of the second order joint probability[7], which is the probability of pixel going gray level i to gray level j with the given distance and direction.

The basic GLCM algorithm

1. Count all pairs of pixels in which the first pixel has a value i, and its matching pair displaced from the first pixel by 'd' has a value of j.
2. This count is entered in the ith row and jth column of the matrix Pd[i,j]
3. Note that Pd[i,j] is not symmetric, since the number of pairs of pixels having gray levels[i,j]does not necessarily equal the number of pixel pairs having gray levels [j,i].
4.The elements of Pd[i,j]can be normalized by dividing each entry by the total number of pixel pairs.
5. Normalized GLCM N[i,j], defined by:

$$N[i,j] = \frac{P[i,j]}{\sum_i \sum_j P[i,j]} \qquad [1]$$

For a window of size wxw, we get one GLCM matrix and the dimension of the co-occurrence matrix is GxG. If we have G gray-levels in the image. Considering distance d and a direction θ Check all pixel pairs with distance d and direction inside the window. Q(i,j|d,θ) is the number of pixel pairs where pixel 1 in the pair has pixel value i and pixel 2 has pixel value j. This has been illustrated in the **Figure- 3(a) and 3(b).**

0	1	1	2	3
0	0	2	3	3
0	1	2	2	3
1	2	3	2	2
2	2	3	3	2

Gray Level j

1	2	1	0
0	1	3	0
0	0	3	5
0	0	2	2

Gray level i

Q(i,j|d,θ).

Fig: 3(a). Image **Fig: 3(b): Pixel Pairs**

....

GLCM features can be used directly to measure statistical measures between the pixels. GLCM extracts the structural information about the texture pattern [8] which has to be analyzed at different orientation and scale. The features which are extracted from GLCM are listed in the **Table- 1**.

Table: 1 . List of GLCM Features

Feature Number	Feature Name	Feature Number	Feature Name
f1	Angular Second Moment (Energy)	f11 [9]	Difference Entropy [9]
f2	Contrast	f12 [9]	Information Measure of Correlation 1[9]
f3	Correlation	f13 [9]	Information Measure of Correlation 2[9]
f4	Sum of Squares: Variance [9]	f14	Autocorrelation
f5	Inverse Difference Moment (Homogeneity) [9]	f15	Dissimilarity
f6	Sum Average [9]	f16	Cluster Shade
f7	Sum Variance [9]	f17	Cluster Prominence
f8	Sum Entropy [9]	f18	Maximum Probability
f9	Entropy	f19	Inverse Difference Normalized
f10	Difference Variance	f20	Inverse Difference Moment Normalized

THE IIOAB JOURNAL ISSN: 0976-3104

Features f1-f13 are features proposed by Haralick [9], Soh proposed features f14-f18 [10] and features f19 and f20 are proposed by Clausi [9].The formula for calculating the some of the features are given below

Energy (angular second moment (asm)

$$f1 = \sum_{i,j=0}^{N-1} P_{i,j}^2 \qquad [2]$$

Contrast

$$f2 = \sum_{i,j=0}^{N-1} P(i,j) * (i-j)^2 \qquad [3]$$

Inverse Difference Moment (IDM) / Homogeneity.

$$f5 = \sum_{i,j=0}^{N-1} \frac{P(i,j)}{1+(i-j)^2} \qquad [4]$$

Entropy

$$f9 = \sum_{i,j=0}^{N-1} P(i,j) * [-\ln(P(i,j))] \qquad [5]$$

Dissimilarity

$$f15 = \sum_{i,j=0}^{N-1} P(i,j) * |(i-j)| \qquad [6]$$

Maximum Probability

$$f18 = \max(i,j)\, P(i,j) \qquad [7]$$

BI-RADS (Breast Imaging Reporting and Data System)

This is developed by American college of Radiology which provides standards for mammographic findings. The standard has been followed by researchers for assessment of different categories of abnormalities present in the mammogram. It specifies the final assessment categories into six categories. It helps in categorising abnormalities as negative, Benign, probably benign, suspicious, malignancy.

METHODOLOGY

For this method, We have taken the image database of digital mammographic images for creating a database of feature extraction from the mini MIAS and it is stored in an excel sheet and loading in matlab. We used the samples of 322 images specified in the Mammographic Image Analysis Society Mini-mammographic Database as our references. In the proposed methodology, the mammogram images are given as input and then noise is removed from the given input image. After removing the noise ,the Otsu method is used for thresholding and then the GLCM features are extracted. The steps which are involved in the methodology is depicted in the **Figure- 3.**

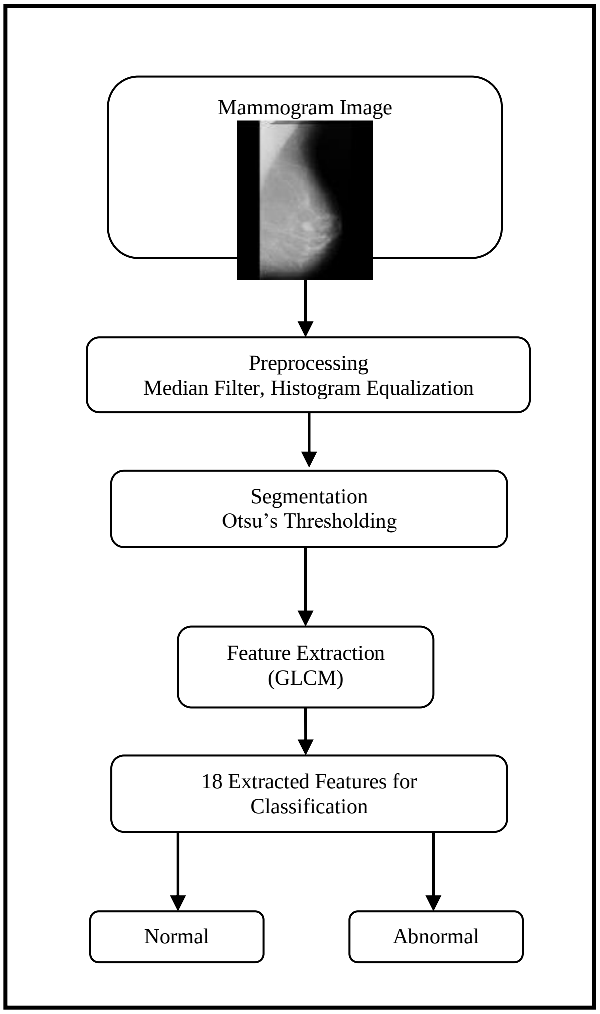

Fig:3. Architecture of the proposed method

………..

Pre-processing

Median Filter is used to remove the noise from the image and improves the quality of the image. Median filter is a well-known order-statistics filter that replaces the original gray value of a pixel by the median of gray values of pixels in the specified neighborhood. A median filter for a smoothed image *f(x,y)* computed from the acquired image *g(x,y)* is defined as

$$f(x,y) = \underset{(i,j) \in N}{Median}\{g(x,y)\} \quad [8]$$

where N is the pre-specified neighborhood of the pixel(x,y) [8].

THE IIOAB JOURNAL ISSN: 0976-3104

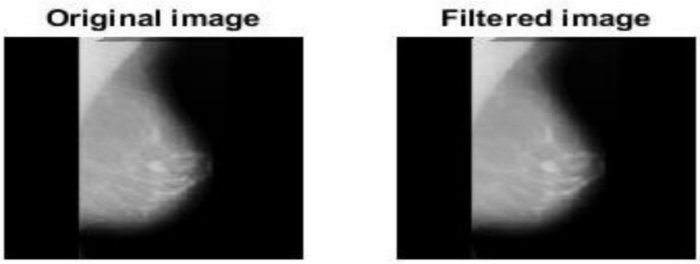

Fig: 4. Original image and output of median filter

……..

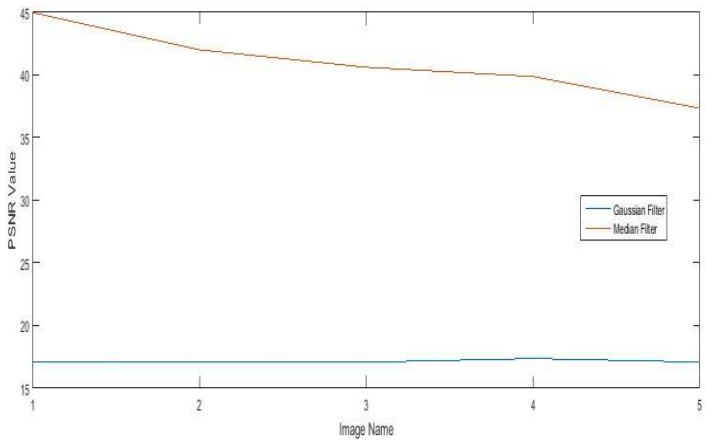

Fig: 5.PSNR value of Gaussian and Median filter

……..

Image Enhancement

The popular technique for enhancing an image is histogram equalization. It is used to reduce the overhead darkness or brightness. It improves the distinct features and visual appearance of the images. The fig.6 shows the histogram of the original image and histogram of the gaussian filtered image.

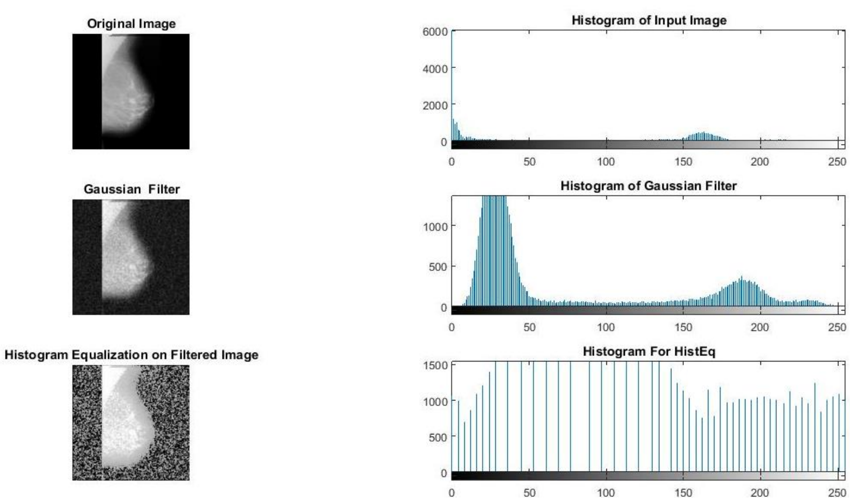

Fig: 6. Histogram of original, Gaussian and histogram equalization

……

THE IIOAB JOURNAL
ISSN: 0976-3104

Median filter gives better result which is shown in the **Figure-,4** .The plot which has been made by considering the PSNR value of different input images shows that the median filters suits well.Median filter removes the noise present in the mammogram and histogram equalization applied for enhancing the input image. It is very clearly observed that the histogram of histogram equalization produces better result.

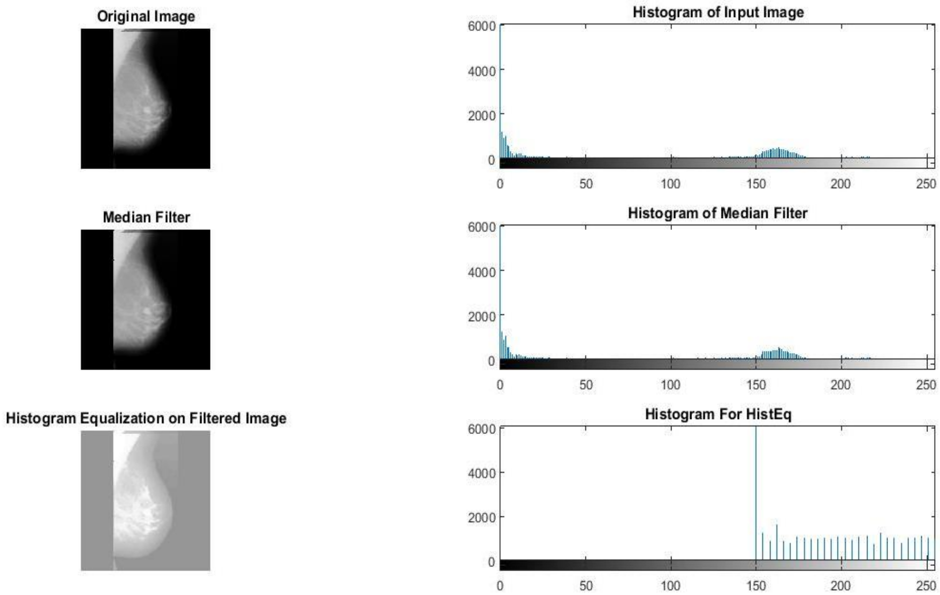

Fig:7. Histogram of median filter and histogram equalization

………

Otsu’s method is used for thresholding the image, which uses global image threshold . graythresh uses Otsu's method, which chooses the threshold to minimize the intraclass variance of the threshold black and white pixels.

Feature Extraction

The GLCM feature algorithm is used to extract features from the result image. Resulting, 18 features of GLCM with each min and max value in the Array. In the **Figure- 5** the output of the input images given to the system which is applied through the median filter is shown in the **Figure- 5**.The output of the filter is further applied with thresholding and the result of this is used for extracting features from GLCM. The features which are extracted from GLCM are tabulated in the **Tables- 2,3,4,5,6 and 7**.

Table: 2. GLCM-Features(From 1-3)

Name of file	Autocorr_1	Autocorr_2	Contrast_1	Contrast_2	corrm_1	corrm_2
216	12.94617293	12.98222935	0.706981897	0.646053619	0.931484	0.937414
10	3.393753944	3.436094222	0.677105823	0.581530948	0.678767	0.723921
141	3.843938419	3.904111704	0.686293956	0.56925649	0.62367	0.687866
32	14.45759369	14.54969316	1.736042078	1.518416825	0.701693	0.739658
248	3.865650149	3.916060406	0.761368647	0.652453979	0.561842	0.624272

ISSN: 0976-3104

Table: 3. GLCM- Features (From 4-6)

Name of file	corrp_1	corrp_2	cprom_1	cprom_2	cshad_1	cshad_2
216	0.931483795	0.937414114	673.8898109	679.6966109	56.38696	56.85415
10	0.678767483	0.723921425	53.87342239	58.64084553	9.695674	10.35827
141	0.623670315	0.687865538	26.75795502	29.80665505	4.511493	4.999365
32	0.701692519	0.739657659	194.0906175	200.3930544	-4.23991	-3.9579
248	0.561842143	0.624272177	20.49201804	23.07003556	3.262367	3.708918

Table: 4. GLCM-Features(From 7-9)

Name of file	dissi_1	dissi_2	energ_1	energ_2	entro_1	entro_2
216	0.385958567	0.3556594	0.273878337	0.277087251	2.195317	2.159219
10	0.413554157	0.37036945	0.343597444	0.353335353	1.848073	1.806057
141	0.471271691	0.41134447	0.197939702	0.208040654	2.146046	2.08653
32	0.835936884	0.74381716	0.059125974	0.064984659	3.168805	3.083147
248	0.52394995	0.46956983	0.174258516	0.18237613	2.20637	2.159627

Table : 5. GLCM- Features(From 10-12)

Name of file	homom_1	homom_2	homop_1	homop_2	maxpr_1	maxpr_2
216	0.847930617	0.85909676	0.837456827	0.849757499	0.506535	0.50854
10	0.830112416	0.84493377	0.818897739	0.835503907	0.572579	0.58112
141	0.796612119	0.81825086	0.78564461	0.809973817	0.394088	0.403445
32	0.689862293	0.71939011	0.664798715	0.698683188	0.136923	0.14237
248	0.773026178	0.79233607	0.761385847	0.783207694	0.353372	0.361869

Table: 6. GLCM- Features(From 13-15)

Name of file	sosvh_1	sosvh_2	savgh_1	savgh_2	svarh_1	svarh_2
216	13.21142782	13.2196865	5.70628882	5.707517566	34.60879	34.74806
10	3.687030929	3.68405151	3.273163218	3.270283066	6.813117	6.925269
141	4.137673514	4.12850937	3.619543657	3.620424506	6.724814	6.852368
32	15.24063245	15.2085565	7.047245233	7.040659239	31.49719	31.67999
248	4.200287118	4.19682665	3.675614596	3.67372823	6.631532	6.714841

Table: 7. GLCM- Features(From 16-18)

Name of file	senth_1	senth_2	dvarh_1	dvarh_2	denth_1	denth_2
216	1.86709554	1.874989983	0.646053619	0.706981897	0.777347	0.818135
10	1.45531407	1.463579527	0.581530948	0.677105823	0.797139	0.851445
141	1.67771022	1.679475856	0.56925649	0.686293956	0.828385	0.896013
32	2.40021975	2.400263073	1.518416825	1.736042078	1.160029	1.225974
248	1.70029865	1.69629512	0.652453979	0.761368647	0.880817	0.937263

In the **Tables** from **2 to 7** ,the extracted 18 features are tabulated and each feature has two ranges which are numbered as 1 and 2.1 indicates the lower range and the 2 indicates the higher range. The name of the features given in the tables represented as

THE IIOAB JOURNAL ISSN: 0976-3104

Autocorr_one- Autocorrelation , Contrast_one- Contrast , corrm_1- Correlation , corrp_1- Correlation, cprom_1- Correlation, cshad_1- Cluster Shade , dissi_1- Dissimilarity , energ_1- Energy , entro_1- Entropy , homom_1- Homogeneity , homop_1- Homogeneity , maxpr_1- Maximum probability , sosvh_1- Sum of sqaures , savgh_1- Sum average , svarh_1- Sum variance , senth_1- Sum entropy , dvarh_1- Difference variance , denth_1- Difference entropy .The features which are extracted from GLCM can be reduced and then the reduced number of features can be given to classifier to classify the abnormalities as benign and malignant.

CONCLUSION AND FUTURE WORK

In this paper, we have extracted GLCM features from the mammogram which has to be done after preprocessing and the segmentation process. The features which are extracted can be used for the further classification technique. The preprocessing is done using median filter and enhancement done through histogram equalization to make the image suitable for segmentation [12] The different features which are extracted based on GLCM can be tried with different classifiers to categorize more precisely the abnormality as normal (benign) and cancerous (malignant).

FINANCIAL DISCLOSURE

No financial support was received to carry out this project.

ACKNOWLEDGEMENT

This work is part of Ph. D Research work. It is not supported by any agency

CONFLICT OF INTERESTS

Authors declare no conflict of interest.

REFERENCES

[1] NCI Cancer Fact Sheets. [Online].Available:http://www.cancer.gov/cancertopics/types/breast

[2] http://www.breastcancer.org/symptoms/testing/types/mammograms/benefits_risks

[3] Kamalakannan J, Tamilarasi Thirumal, Abinaya Vaidhyanathan, and Kansagara Deep MukeshBhai.[2015.] Study on different classification technique for mammogram image", 2015International Conference on Circuits Power and Computing Technologies [ICCPCT-2015],

[4] Kamalakannan J, P Venkata Krishna, M. Rajashekara Babu, and Kansagra DeepMukeshbhai.[2015] Identification of abnormility from digital mammogram to detect breast cancerInternational Conference on Circuits Power and Computing Technologies [ICCPCT- 2015], 2015

[5] RM Haralick, K Shanmugam, and I Dinstein.[1973] Textural features for image classification, IEEE Transactions on systems, man and cybernetics, 3(6): 610-621.

[6] DA Clausi.[2002] An analysis of co-occurrence texture statistics as a function of grey level quantization, *Canadian Journal of Remote Sensing*, 28(1): 45–62.

[7] Kulkarni, Nilambari, and Vanita Mane. [2015] Sourcecamera identification using GLCM", 2015 IEEE International Advance Computing Conference (IACC), 2015.Kamalakannan, J., Tamilarasi Thirumal, Abinaya Vaidhyanathan, and Kansagara Deep MukeshBhai. "Study on different classification technique for mammogram image", 2015International Conference on Circuits Power and Computing Technologies [ICCPCT-2015].

[8] Saroja G Arockia Selva, C Helen Sulochana.[2013] Texture analysis of non-uniform images using GLCM, 2013 IEEE Conference on Information and Communication Technologies.

[9] Radovic, Milos, Marina Djokovic, Aleksandar Peulic, and Nenad Filipovic. "Application ofdata mining algorithms for mammogram classification", 13th IEEE International Conference on BioInformatics and BioEngineering, 2013.

[10] LK Soh, and C Tsatsoulis.[1999] Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices,*" IEEE Transactions on geoscience and remote sensing*, 37(2): 780-795,

[11] Sameer Singh, and Keir Bovis.[2005] An Evaluation of Contrast Enhancement For breast Techniques for Mammographic Breast Masses, *IEEE Transactions On Information Technology In Biomedicine*, 9(1): 109-119

[12] Jawad Nagi, Sameem Abdul Kareem, Farrukh Nagi, Syed Khaleel Ahmed.[2010] Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms, *IEEE EMBS Conference on Biomedical Engineering & Sciences*

[13] H Abdellatif, TE Taha, OF Zahran, W Al-Nauimy, FE. Abd El-Samie.[2013] K9. Automatic Segmentation of Digital Mammograms to Detect Masses, *IEEE* .

[14] Tiago T Wirtti, Evandro OT. Salles. Segmentation of Masses in Digital Mammograms, *IEEE* pp 1-6.

[15] A Oliver, J Freixenet ,J Marti,Elsa Perez, J Pont and RE.Denton.[2010] A review of Automatic mass Detection and Segmentation in Mammographic Images, Medical Image Analysis, 14-2 :.87-110

[16] D.Brzakovic et.al.[1990] An approach to automated detection of tumors in mammograms, *Medical Imaging, IEEE Transactions on*, 9:233-241.

[17] RM Haralick, K Shanmugam, I Dinstein.[1973]Textural features for image classification, *IEEE Transactions on systems, man and cybernetics*, 3(6):610-621.

THE IIOAB JOURNAL
ISSN: 0976-3104

[18] DA Clausi.[2002] An analysis of co-occurrence texture statistics as a function of grey level quantization, *Canadian Journal of Remote Sensing*, 28(1): 45–62.

[19] LK Soh, and C Tsatsoulis.[1999] Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices," *IEEE Transactions on geoscience and remote sensing*, 37(2): 780-795.

[20] Kamalakannan J, P Venkata Krishna, M Rajashekara Babu, and Kansagra DeepMukeshbhai.[2015] Identification of abnormility from digital mammogram to detect breast cancer",2015 International Conference on Circuits Power and Computing Technologies [ICCPCT- 2015]

[21] RM Haralick, K. Shanmugam, I Dinstein. [1973] Textural Features for Image Classification, IEEETransactions on Systems, *Man and Cybernetics*, 3(6): 610–621.

[22] FI Alam, RU Faruqui. [2011] Optimized Calculations of Haralick Texture Features, European Journal of Scientific Research, 50 (4): 543-553.

[23] D A. Clausi, (2002) An analysis of co-occurrence texture statistics as a function of grey level quantization, Can. J. Remote Sensing, Vol. 28, No. 1, pp 45-62.

[24] Radovic, Milos, Marina Djokovic, Aleksandar Peulic, and Nenad Filipovic.[2013] Application ofdata mining algorithms for mammogram classification", 13th IEEE International Conference on BioInformatics and BioEngineering.

[25] Kulkarni, Nilambari, and Vanita Mane.[2015] Sourcecamera identification using GLCM", 2015 IEEE International Advance Computing Conference (IACC),.

[26] elena Bozek, Mario Mustra, Kresimir Delac, and Mislav Grgic."A Survey of Image Processing Algorithms in Digital Mammography, *J Bozek* et al. 635.

[27] Jain AK, Duin RPW, Mao J.[2000] Statistical Pattern Recognition, A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1), 4–37

ABOUT AUTHORS

Prof.Kamalakannan.J, He is a faculty member at School of Information Technology and Engineering, VIT University, Vellore, India. He has completed his Bachelors in Electronics and Communication Engineering and Masters in Computer Science and Engineering from Madras University, Currently pursuing research at School of Computing Science and Engineering,VIT University, India

Dr.M.Rajasekhara Babu, He is a Senior faculty member at School of Computer Science and Engineering, VIT University, Vellore, India. He has completed his Bachelors in Electronics and Communication Engineering from Sri Venkateswara University, Tirupathi, India and took his Masters in Computer Science and Engineering from Regional Engineering College(NIT), Calicut and he has completed his Ph.D from VIT University,India.