

Original Research Article

Detection of Tumour Based on Breast Tissue Categorization

ABSTRACT

Background: Despite the benefits of Computer Aided Detection (CAD), false detection of breast tumour is still a challenging issue with oncologist. A mammography is a non-invasive screening tool that uses low energy X-rays to show the pathology structure of breast tissue. Interpreting mammogram visually is a time consuming process and requires a great deal of skill and experience. Earlier Computer Aided Techniques emphasis detection of tumour in breast tissues rather than categorization of breast into Breast Imaging Report and Data System (BI-RADS) which is the medically understandable method of reporting.

Aim: The work centered on developing a CAD system which is capable of not only detecting but also categorizing breast tissue in line with BI-RADS scale.

Methodology: The acquired images were pre-processed to remove unwanted contents. Two stage medical procedural approach was designed to categorize the tissue in breast images into low dense (fatty) and high dense. Tumours in the low dense breasts were segmented, and then classified into normal, benign and malignant. The developed system was evaluated using sensitivity, specificity, false positive reduction, false negative reduction and overall performance.

Results: The developed CAD achieved 90.65% sensitivity, 73.59% specificity, 0.02 positive reduction, 0.04 false negative reduction and 85.71% overall performance.

Conclusion: The false positive reduction result obtained shows that false detection has been minimized as a result of categorization procedure of the breast tissue in mammograms.

Keywords: Bi-lateral, BI-RADS, Categorization, Computer Aided Detection, Pixels, Pre-processing, Segmentation, Tumour

1. INTRODUCTION

Breast cancer originates in breast tissue, which is made up of glands for milk production (lobules), and the ducts that connect lobules to the nipple [3]. Breasts contain both dense tissue (glandular tissue and connective tissue, together known as fibro-glandular tissue) and fatty tissue. Fatty tissue appears dark on a mammogram, whereas fibro-glandular tissue appears as white. According to [11], it is harder to detect tumours in women with denser breasts because fibro-glandular tissue and tumours have similar density.

The American College of radiology introduced Breast Imaging Reporting and Data System (BI-RADS) categorization. The four categories of the density system are: BI-RADS1 represents predominantly fatty breast; BI-RADS2-scattered fibro-glandular densities; BI-RADS3-breast that is heterogeneously dense and BI-RADS4-highest level, an extremely dense breast that could obscure a lesion [19]. The breast images that represent each of the breast tissue categorization are shown in Fig. 1.

Different image modalities are used in the detection and evaluation of breast abnormalities among such is mammography. A mammography is a non-invasive screening tool recommended for young women who have symptoms of breast cancer or have a high risk of breast cancer, as well as for women older than 40 years even if there is no sign of the disease [6]. In spite of the benefits of mammography in detection of abnormal cells in the breast, visual interpretation of mammogram is

a time consuming process and requires a great deal of skill and experience. However, image interpretation can be improved using computational advancement [10]. Therefore, there is need for better, inexpensive and automatic methods for cancer detection and evaluation. CAD is a secondary tool design for radiologist to detect breast cancer. The goal of Computer Aided Detection (CAD) is to improve radiologist's performance by indicating the sites of potential abnormalities, to reduce the number of missed abnormality, and by providing quantitative analysis of specific regions in an image to improve diagnosis.

Pattern recognition methods are widely used in computer vision system to analyze and recognize the image content. [17] described the process of pattern recognition and classification as: physical environment, data acquisition (test data), pre-processing (including segmentation), feature extraction, classification, post processing, and decision. This work developed a Computer Aided Detection (CAD) system that analyzed digital mammogram to detect breast tumour and classify mammograms into normal, benign or malignant. Computer vision technique, which involves medical procedural approach, segmentation technique, feature extraction and a classification technique were used in the design of the system.

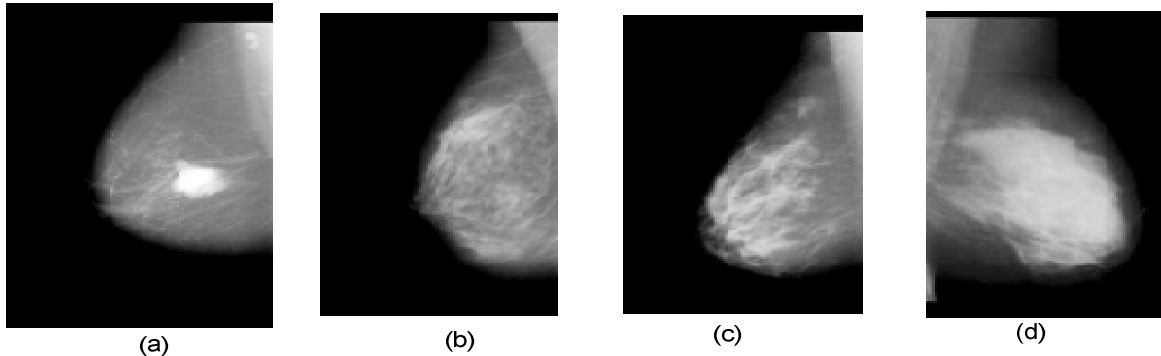


Fig.1. Mammograms with differing mammographic breast densities
(a) predominantly fat; (b) fat with some fibroglandular tissue;
(c) heterogeneously dense; (d) extremely dense [19].

2. RELATED WORKS

There exists numerous research on categorization of mammograms based on breast density. [9] used granulometry and texture energy to classify breast tissue into fatty and glandular breast types. [18]. classified fatty and dense breast types using an automated method of extracting the Region of Interest (ROI) based on texture. [4] analysed two different classification methods, which are four-class categories according to the BI-RADS system and two-class categories, differentiating between dense and fatty breast types. The results showed that the classification based on BI-RADS system for the four-class categories (average recognition rate, 71.4%) is a challenging task in comparison to the two-class categories (average recognition rate, 96.7%). [21] classified breast density into one of four BI-RADS categories according to the characteristic features of gray level histogram and found that the correlation between computer-estimated percent dense area and radiologist manual segmentation was 0.94 and 0.91 with root-mean-square (RMS) errors at 6.1% and 7.2%, respectively, for Cranio-Caudal (CC) and Medio-lateral Oblique (MLO) views.

[8] divided breast mammogram images into three regions using variance histogram analysis and discriminant analysis. Then, classify it into four categories, which are fatty, mammary gland diffuseness, non-uniform high density, and high density, by using the ratios of each of the three

regions. [4] estimate features from constructed Spatial Gray Level Dependency matrices and trained multiple Artificial Neural Networks (ANN) to achieve two-categories (BI-RADS I and II versus BI-RADS III and IV) and four categories parenchymal pattern classification.

[13] used Bayesian combination to categorize mammography Image Analysis Society (MIAS) breast images into dense and fatty breast according to the BIRADS lexicon. [20] used an approach by [14], which adopted a Bayesian combination of decision tree and the k-Nearest Neighbor (KNN) algorithm to classify the breast according to BI-RADS categories. [15] implemented KNN classifier to classify breast tissue into fatty and dense. [4] Classified mammographic breast density using a combined classifier paradigm. [19] presented a tool that detected breast edges, dividing the mammary structure and the background of the image, identify density region and then classified mammograms by density through computation techniques such as local statistics and texture measures to divide the mammograms in fat or dense. [19] shows that in women with density above 75%, the cancer risk increases to four times women with less density.

3. BREAST TISSUE CATEGORIZATION

The complete framework for breast tissue categorization and detection of breast tumour is presented in Fig. 2 while the flow chart of the developed system is shown in Fig. 3.

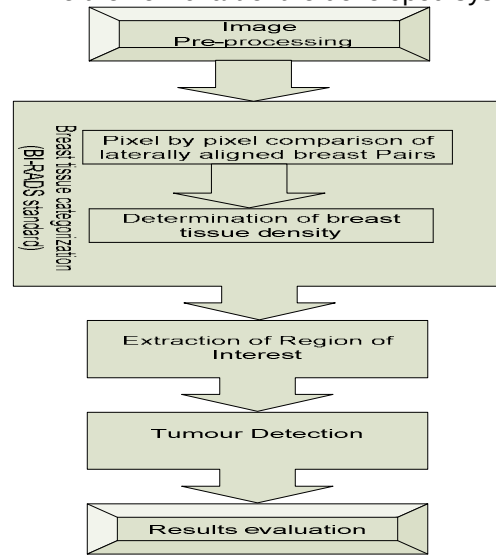


Fig. 2. Framework of the developed system

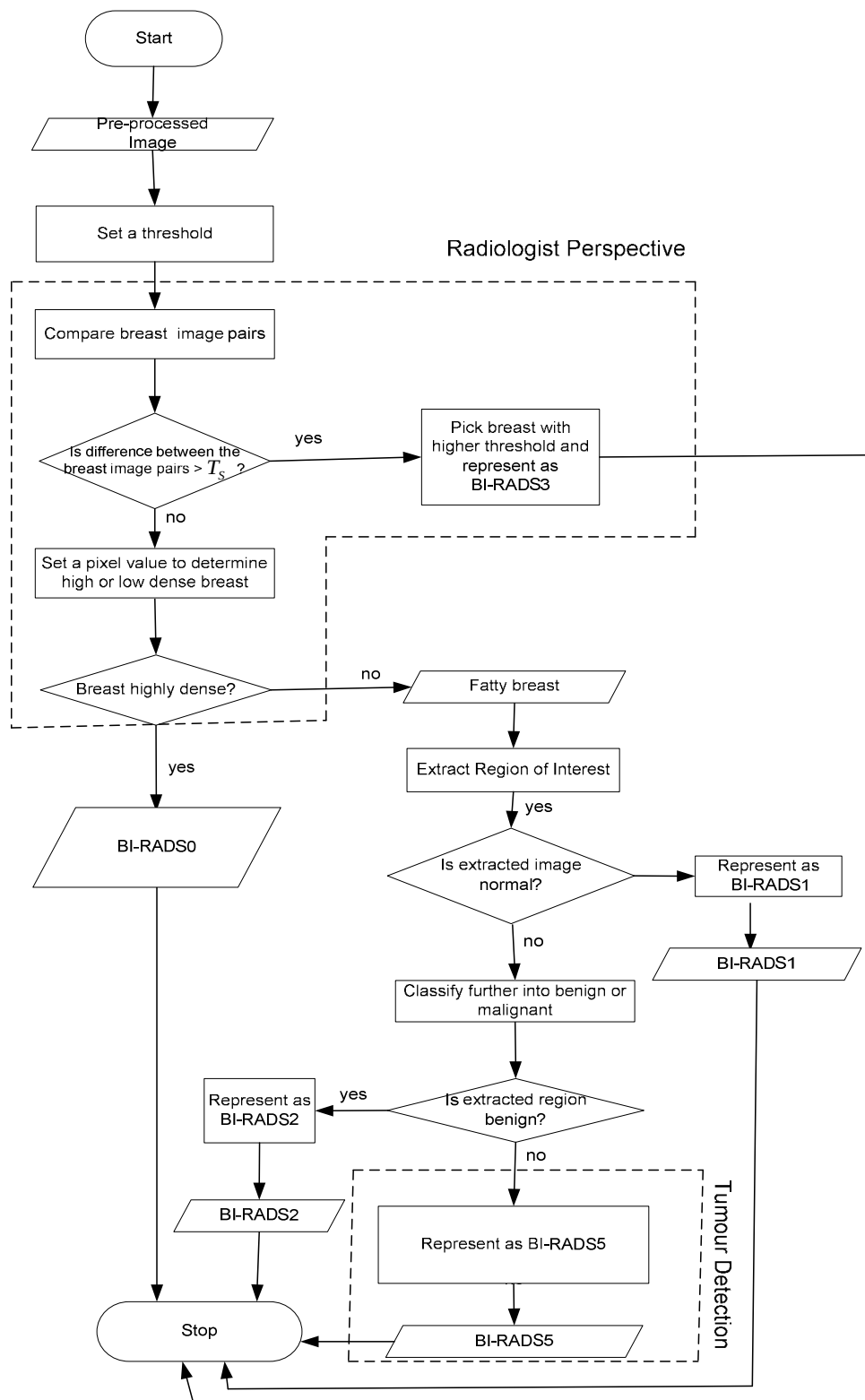


Fig. 3. Flow chart of the developed system

The acquired breast images were pre-processed, in order to remove artifacts and suppress pectoral muscle using regional description technique as explained in [12]. Mammogram density categorization entails bilateral comparison of pairs of breast image to detect suspicious region in the breast image and setting a pixel value (T_s) to 0.1 for the image, which aided in identification of high dense breast. The intensity value is in line with density value used by [1] to characterized breast background tissue. The BI-RADS categorization of breast images was done in line with the American College of Radiology (ACR) recommendation. Region of **interest** is extracted from low dense (fatty) breast images to aid detection of breast tumour.

Breast tissue categorization is essential due to the fact stated by [5] that "Breast area with high dense tissue caused decrease detection of cancers". The breast tissue categorization procedure was done by pixel to pixel comparison of bilaterally rotated and complementary breast pair. The original right and left MLO (Medio-lateral Oblique) images are shown in Fig. 4 (i) and (ii) tagged as Im_R and Im_L respectively. **The right breast image was flipped and tagged as Im_R^1 , in order to have the same orientation with left breast thereby, making the comparison process simpler as shown in Fig. 4 and 5.**

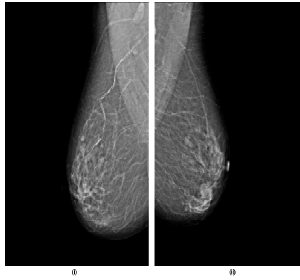


Fig. 4. Original MLO images of the (i) right (ii)left

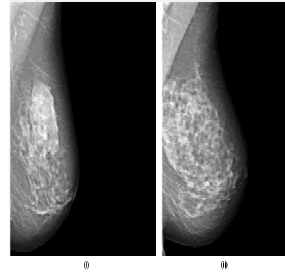


Fig. 5. (i) Flipped right (ii)original left

After flipping, a threshold value of 161 out of 0-255 range in an 8-bits coded gray-scale image was set for high intensity pixel. The pixel intensity score for each was estimated for each pair of the breast image by checking if Im_R^1 and Im_L are greater than the set threshold (161). The absolute difference of Im_R^1 and Im_L was determined using the equation below:

$$\frac{|\sum P(Im_R^1) - \sum P(Im_L)|}{1024^2} \geq 0.05$$

Another threshold value of 0.05 out of 0-1 range in an 8-bits coded gray-scale image was set, to compare the absolute difference value. The breast image with higher pixel intensity score was picked, if the absolute difference is greater than the set threshold (0.05). In a case where there is a sharp difference in the piecewise comparison, the pair with higher value of pixels intensity is picked, suspected to be cancerous and classified to **BI-RADS3**.

Breast (physiological and pathological) pattern identification, the second phase of breast density categorization is necessary in order to see a correlation between breast density and an increase risk of breast cancer. The two approaches of categorizing mammogram into class are four-class and two-class approach. Four class approach are ACR standard category as stated in Table 3, while the two-class approach was derived from the four ACR standard way of categorising breast image as type1 were categorized as low density breast while type2, type3 and type 4 were categorized as high density. These two-class approach from mammographic risk assessment point of view might be more appropriate than four-class division [13].

The developed system followed the existing work which pre-sorted mammograms into "fatty" and "dense" categories; so that the "dense" can be read by a more experienced mammographer. This stage entails using a pixel value to know the characteristic of background

tissue of the images (fatty or dense). An intensity value of pixel was set to 0.1 out of 0-1 range in an 8-bit coded gray-scale image, which was used to determine the percentage of the highly dense breasts. Any image that has a pixel value above this set value was categorised as highly dense breast otherwise they are categorized as low dense breasts. The highly dense breast images were categorized as BI-RADS0 (additional imaging evaluation is required).

In order to detect breast tumour, salient features from the segmented regions following the technique in [2] were put into Multiclass Support Vector Machine (MSVM) to classify images into normal, benign and malignant. The results of the developed CAD system were evaluated using, sensitivity (ability to identify abnormal mammograms), specificity, false positive reduction, false negative reduction and accuracy. The metrics that were used for the evaluation of breast background tissue characteristics are: sensitivity and (SE) specificity (SP) and accuracy as stated below:

$$SE = \frac{TP}{TP + FN} \quad , \quad SP = \frac{TN}{TN + FP} \quad , \quad \text{False positive reduction} = \frac{FP}{\text{Total image}} \quad ,$$

$$\text{False negative reduction} = \frac{FN}{\text{Total image}} \quad \text{and} \quad \text{Accuracy} = \frac{TP + TN}{\text{Total number of image}}$$

4. ALGORITHM FOR BREAST IMAGES BILATERAL COMPARISON

The algorithm for the bilateral comparison of breast images of the same object is given below. The algorithm was experimented on the pre-processed breast images.

```
% to put both right and left breast in the same direction
let represents right breast
let represents left breast
let represents flipped right breast
set threshold value of 161 out of 0-255 range in an 8-bits
coded gray-scale image
% to estimate pixel intensity score of both flipped right and
left breast
if
if

% to identify significant difference between the flipped right
and left breast

% to identify suspicious breast
if
pick image with higher
then the image is suspected and put as BI-RADS3
```

5. RESULTS AND DISCUSSION

The result obtained at the first stage of the breast tissue categorization shows that 160 out of 320 pre-processed images were suspected of abnormality. It was observed from Table 1 that at

ACR type1, specificity is 73.59%; at ACR type2, specificity is 90.57%; at ACR type3, specificity is 96.23%; and at ACR type4 specificity 98.11%. The specificities at ACR type2, ACR type3, and ACR type4 indicate that it will be difficult to identify abnormality in the mammogram which is in line with [5] that 'women with more than 75% density had an increase risk of breast cancer', therefore another imaging evaluation is required. Sensitivity was used to evaluate the pathological pattern of mammograms. It was observed from Table 1 that at ACR type1, sensitivity is 90.65%; at ACR type2, sensitivity is 61.68%; at ACR type3, sensitivity is 20.56%; and at ACR type4, sensitivity is 5.61%. The sensitivity (90.65%), specificity (73.59%) and overall accuracy (85%) at ACR type1 show that abnormality can easily be identified with fatty tissue in the mammograms at higher rate than other ACR type in the category.

Table 1. The results of ACR breast tissue types

Metrics	Results	ACR Type	Density Percentage Value (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
TP	97	1 (fatty breast)	<10	90.65	73.59	85.00
FP	14					
TN	39					
FN	10					
TP	66	2 (Fibro-glandular dense)	25-50	61.68	90.57	71.25
FP	5					
TN	48					
FN	41					
TP	22	3 (Heterogeneous dense)	50-75	20.56	96.23	45.63
FP	2					
TN	51					
FN	85					
TP	6	4 (Extremely dense)	75>	5.61	98.11	36.25
FP	1					
TN	52					
FN	101					

The algorithm step through different threshold values out of 0-255 range in an 8-bits coded gray-scale image to ascertain better density value for the fatty breast. This is shown in Table 2. It was observed that the sensitivity at threshold value 171 is 90.65%. The sensitivity at threshold values 173, 175, 177, 179 and 181 are 89.72%, 89.72%, 84.11%, 81.31% and 77.57% respectively. According to [7], the larger number of false positive and fewer number of false negative confirmed presence of cancer, further investigation must be undertaken. It was observed from Table 2 that at threshold value 171, the larger number of false positive and the fewer the number of false negative compare to other threshold values show that majority of the images negated by the algorithm are fatty breasts, therefore identification of abnormality will be easier.

Table 2. Results of breast tissue characterization at different threshold value

Metrics	Results	Threshold Value	Sensitivity (%)	Accuracy (%)
TP	97	171	90.65	85.00
FP	14			
TN	39			
FN	10			
TP	96	173	89.72	
FP	9			

TN	44			87.50
FN	11			
TP	96			
FP	9	175	89.72	87.50
TN	44			
FN	11			
TP	90			
FP	8	177	84.11	84.38
TN	45			
FN	17			
TP	87			
FP	7	179	81.31	83.13
TN	46			
FN	20			
TP	83			
FP	5	181	77.57	81.88
TN	48			
FN	24			

Some of the output images of the highly density score were shown in Fig. 6. The patients that are in this category will need other screening method in future, in order to clarify that there is no abnormality.

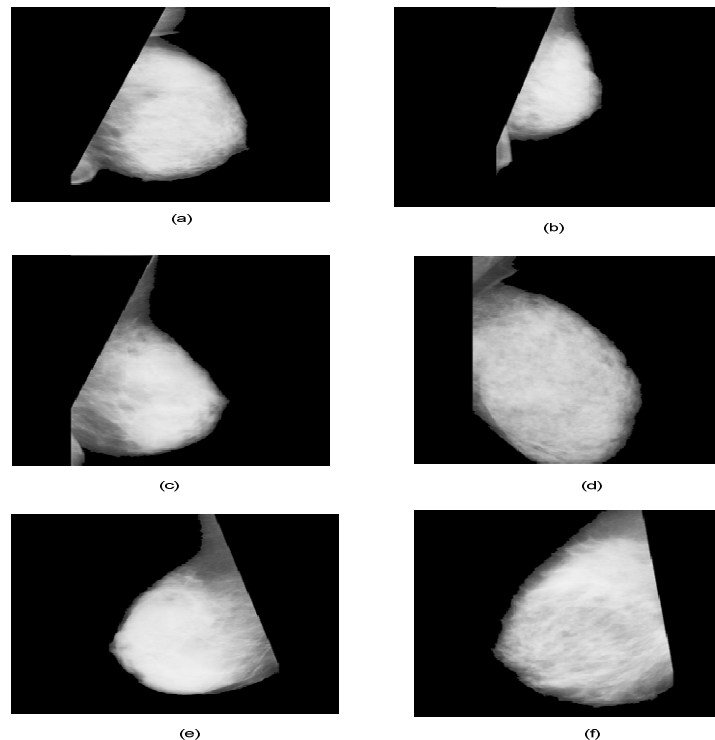


Fig. 6. high dense mammogram at density value greater than 10%

The bar chart in Fig. 7 shows the sensitivity and accuracy at different threshold values, of the breast density categorization. The sensitivity at threshold 171 is higher than other threshold in the table, which ascertained that at threshold value of 171 of ACR type1 mammograms, abnormality can be identified more clearly.

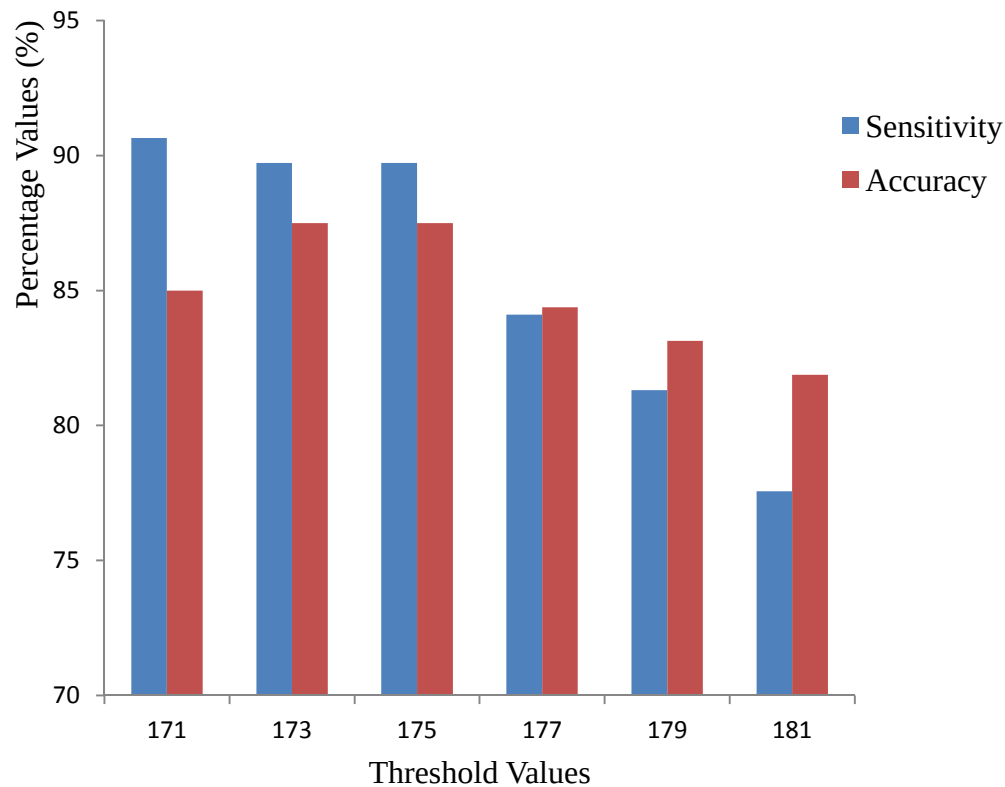


Fig. 7. Bar chart showing breast tissue categorization at different threshold values

The accuracy of breast density categorization obtained from the developed system is compared with existing similar works as shown in Table 3. It was observed that the accuracy (85%) of the proposed using two-class approach is better than [4] accuracy (71%) and [16] accuracy (76%) that used four-class approach of combination. The two existing works mentioned, employed classifier-based method. [13] using four-class approach obtained 86% accuracy but used another database Digital Database for Screening Mammography (DDSM) from the University of South Florida, while the method was texture-base. The developed system followed two-class approach of combination because, only fatty breast make the abnormality detection easier. Also, the mammographic risk assessment point of view will be more appropriate in the developed system than the existing four-class approach. If other ACR types of breast tissue is combined with the ACR type1 (fatty), there may be an increase in false positive, causing mislabel of affected patient as normal patient.

Table 3: Comparison of breast tissue categorization based on two-class approach result with four class approach existing works

Authour	Method	Database Used	Class Combination Approach	Categorization Accuracy
[4]	Classifier-based	MIAS	Four-class	71%

[16]	Classifier-based	MIAS	Four-class	76%
[13]	Texture-based	DDSM	Four-class	86%
Developed system	Pixel-based	MIAS	Two-class	85%

Two class approach: fatty breast=low density
fibro-glandular, heterogeneously dense, homogeneously dense=high density

The detection of breast tumour employed classification procedure of MSVM (Multiclass Support Vector Machine) classifier in MATLAB version 11. Out of 49 images classified, there were 28 "normal", 12 "malignant", and 9 "benign". The results obtained from classification stage is shown in Table 4. It was observed at threshold value of 0.900, 27 were correctly classified as "normal", while 1 was misclassified as "malignant" (false positive), 10 were correctly classified as "malignant", 2 were wrongly classified as "normal" (false negative), 5 were correctly classified as "benign" while 4 were wrongly classified as "normal". The false positive reduction and false negative reduction at threshold of 0.900 were 0.02 and 0.04 respectively. The same was done for other threshold values.

The sensitivity of the classifier to determine whether the abnormality was malignancy or not at threshold value (0.90) was 83% and the classification accuracy was 85.75%. It was observed that the sensitivity at threshold values (0.900), (0.915) and (0.925) was 83% but the overall accuracy of the classifier at 0.900 was 85.71%, which is higher than 79.59% and 75.51% of threshold values (0.915) and (0.925) respectively. Also, the sensitivity at threshold values (0.945), (0.950) were (58%) and (42%) respectively. Although, the false positive reduction at threshold values (0.900) and (0.945) were the same but the overall performance of the classifier at threshold value (0.900) was higher than at threshold values (0.945). The overall performance of the classifier at threshold value (0.900) is the best compared with other threshold values in the table.

Normal mammograms were categorized as BI-RADS1, benign form of tumour were categorized as BI-RADS2 while malignant form of tumour were categorized as BI-RADS5, using BI-RADS assessment. Some of the results of the classified mammograms are shown in Figure 8(a-c). The Graphic User Interface (GUIs) show example of normal mammogram, benign and malignant. Also, the GUIs indicate the reference result to benchmark the result of the developed system. It can be observed from the GUIs that the mammograms are correctly classified by the MSVM.

Table 4. Overall performance of multi-SVM classifier using confusion matrix

Thres hold Value	Actual Class				Metrics	Result s	Sensitivit y	False Positive Reductio n	False Negative Reductio n	Overall Classifier Performance
	Normal	Malignant	Benign							
0.900	N	27	2	4						
	M	1	10	0	TP	10	83.00%	0.02	0.04	85.71 %
					FP	1				
	B	0	0	5	TN	36				
					FN	2				
0.915	N	25	2	5						
	M	2	10	0	TP	10	83.00%	0.04	0.04	79.59 %
					FP	2				
	B	1	0	4	TN	35				
					FN	2				

0.925	N	25	2	6							
	M	3	10	1	TP	10					
					FP	4	83.00%	0.068	0.04	75.51	
	B	0	0	2	TN	33				%	
					FN	2					
0.935	N	25	3	6							
	M	3	9	0	TP	9					
					FP	3	57.00%	0.06	0.06	75.51	
	B	0	0	3	TN	34				%	
					FN	3					
0.945	N	27	5	9							
	M	1	7	0	TP	7					
					FP	1	33.00%	0.02	0.10	69.39	
	B	0	0	0	TN	36				%	
					FN	5					
0.950	N	27	6	6							
	M	1	5	1	TP	5					
					FP	2	33.00%	0.04	0.14	69.39	
	B	0	1	2	TN	35				%	
					FN	7					

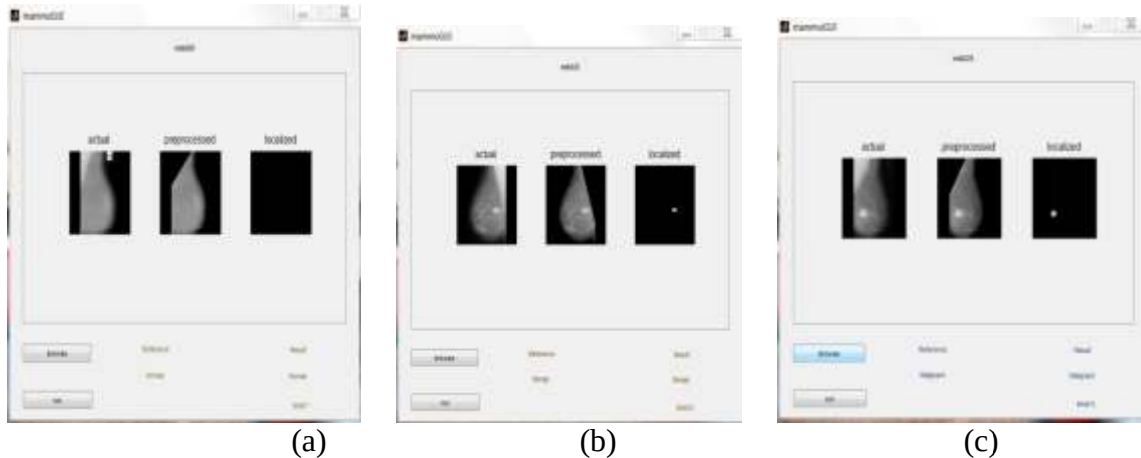


Fig. 8.: Matlab GUI for (a) normal (b) benign (c) malignant

5. CONCLUSION

Medical procedural approach in the developed system assisted in identification of suspicious mammograms and identification of dense and fatty breasts. Detection of tumour is easy low dense (fatty) breast than highly dense breast. Finally, the classification of the segmented mammogram into normal, benign and malignant achieved a better false positive reduction (0.02) and false negative reduction (0.04). The developed CAD system has improved the methods of detection and classification of breast tumour in terms of overall performance.

REFERENCES

1. Adepoju T. M, (2015). " development of a breast cancer detection and classification system using two-stage segmentation technique", A PhD thesis submitted to the Department of Computer Science and Engineering, LAUTECH, Ogbomosho, Nigeria.
2. Adepoju T. M., Ojo J. A., Omidiora E. O., and Olabiyisi S. O. (2015). "Pixel-Based Morphological Technique for Breast Tumour Detection". International Journal of Scientific & Engineering Research, Volume 6, Issue 6, pp. 1416-1420.
3. Bandyopadhyay, S. K., Maitra I. K. and Banerjee, S. (2010). "Digital imaging in pathology towards detection and analysis of human breast cancer", Second International Conference on Computational Intelligence, Communication System and Networks, IEE Computer Society, Vol. 43, p. 295.
4. Bovis, K. and Singh, S. (2002) "Classification of mammographic breast density using a combined classifier paradigm", in Proc. Med. Image Understanding Anal. Conf., pp.177–180.
5. Boyd N. F., Huszti E., Melnichouk O., Martin L. J., Hislop G., Chiarelli A., Yaffe M. J. and Minkin S. (2014). "Mammographic features associated with interval breast cancers in screening programs", Breast Cancer Research, Vol. 16, Issue 417, pp.1-9.
6. Eden, A. A., Jose, L. H., Jesus, S. C., Camila, C., Hugo, T., Santiago, E. C. (2012). "Analysis of Machine Learning Techniques Applied to the Classification of masses and Microcalcification Clusters in Breast Cancer Computer-Aided Detection", *Journal of Cancer Therapy, Scientific Research*, pp. 1020-1028.
7. Mangrulkar, R. (2012). "[A lecture on Diagnostic Reasoning](http://open.umich.edu/education/med/m1/patientspop-decisionmaking/2010/materials)", M1 - Patients and Populations Sequence, the University of Michigan Medical School's. <http://open.umich.edu/education/med/m1/patientspop-decisionmaking/2010/materials>
8. Matsubara, T., Yamazaki, D., Kato, M., Hara, T., Fujita, H., Iwase, T. and Endo, T. (2001) "An automated classification scheme for mammograms based on amount and distribution of fibroglandular breast tissue density", International Congress, series 1230.
9. Miller, P. and Astley, S. M. (1992) "Classification of breast tissue by texture analysis", Image Vision Comput., Vol. 10, pp. 277–282.
10. Mishra, Suman and Ranganathan Hariharan (2013). 'Dyadic Wavelet Transform based Classification of Microcalcifications Using SVM', Journal of Computer Applications (JCA), Vol. VI, Issue 3, pp. 51-56.
11. National Cancer Institute, (2012). "*fact sheet*". Department of Health and Human Services National Institute of Health, US.
12. Ojo J. A., Adepoju T. M., Omidiora E. O., Olabiyisi O. S. and Bello O. T, (2014) "Pre-Processing Method for Extraction of Pectoral Muscle and Removal of Artefacts in Mammogram", *IOSR journal of Computer Engineering*, Vol.16, Issue 3, pp 06-09. e-ISSN: 2278-0661, p- ISSN: 2278-8727.
13. OliverMalagelada, O. A. (2007). "*Automatic Mass Segmentation in Mammographic Images*", PHD Thesis, University of Girona, Girona.
14. Oliver, A., Freixenet, J., Martí, R., Pont, J., Pérez, E., Denton, E. R. and Zwigglelaar, R. (2008). "A Novel Breast Tissue Density Classification Methodology", IEEE Tran. Inf Technology Biomed, Vol. 12, No. 1, pp. 55-65.
15. Oliver, A., Lladó, X., Pérez, E., Pont, J., Denton, E., Freixenet, J. and J. Martí, (2009) "A Statistical Approach for Breast Density Segmentation", Journal of Digital Imaging, pp. 1-11.

- 359 16. Petroudi S. Kadir T. and Brady M. (2003). "Automatic classification of mammographic
360 parenchymal patterns: A statistical approach", *In IEEE Conf. Eng. Med. Biol. Soc.*, Vol. 2,
361 pp. 416-423.
- 362 17. Sharma, Priyanka and Kaur, Manavjeet (2013). 'Classification in Pattern Recognition: A
363 Review', *International Journal of Advanced Research in Computer Science and Software*
364 *Engineering*, Vol. 3, Issue 4, pp. 298-306.
- 365 18. Taylor, P., Hajnal, S. Dilhuydy, M-H. and Barreau, B. (1994). "Measuring image texture to
366 separate difficult from easy mammograms", *The British Journal of Radiology*, Vol. 67, pp.
367 456-463.
- 368 19. Teixeira Rita (2013). "Automatic Analysis of Mammography Images: Classification of
369 Breast Density", MSc thesis in Biomedical Engineering, Faculdade de Engenharia da
370 Universidade do Porto, pp. 15-17.
- 371 20. Torrent, A., Bardera, A., Oliver, A., Freixenet, J., Boada, I., Feix, M., and Martí, J. (2008)
372 "Breast Density Segmentation: A Comparison of Clustering and Region Based
373 Techniques", (eds.) IWDM. LNCS, Springer, Heidelberg, Vol. 5116, 2008, pp. 9-16.
- 374 21. Zhou, C., Chan, H. P., Petrick, N., Helvie, M. A., Goodsitt, M. M., Sahiner, B. and
375 Hadjiiski, L. M. (2001), "Computerized image analysis: estimation of breast density on
376 mammograms", *Medical Physics*, Vol. 28, No. 6, pp. 1056-1069.