

Automated Classification of Breast Cancer Lesions for Digitised Mammograms via Computer-Aided Diagnosis System

Saifullah Harith Suradi^{a,*}, Kamarul Amin Abdullah^{a,b} and Nor Ashidi Mat Isa^c

^a*School of Medical Imaging, Faculty of Health Sciences, Universiti Sultan Zainal Abidin (UniSZA), 21300 Kuala Nerus, Terengganu, Malaysia*

^b*Medical Imaging Research Group (MIRG), Faculty of Health Sciences, Universiti Sultan Zainal Abidin (UniSZA), 21300 Kuala Nerus, Terengganu, Malaysia*

^c*School of Electrical and Electronic Engineering, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia*

Abstract

Women with breast cancer have a high risk of death. Digitised mammograms can be used to detect the early stage of breast cancer. However, digitised mammograms suffer low contrast appearances that may lead to misdiagnosis. This paper proposes a Computer-Aided Diagnosis (CAD) system of automated classification of breast cancer lesions using a modified image processing technique of Fuzzy Anisotropic Diffusion Histogram Equalization Contrast Adaptive Limited (FADHECAL) incorporated with Multilevel Otsu Thresholding on digitised mammograms. Four main blocks were used in this CAD system, namely; (i) Pre-processing and Enhancement block; (ii) Segmentation block; (iii) Region of Interests (ROIs) Extraction block; and (iv) Classification block. The CAD system was tested on 30 digitised mammograms retrieved from the Mini-Mammographic Image Analysis Society (MIAS) database with various degrees of severity and background tissues. The proposed CAD system showed a high accuracy of 96.67% for the detection of breast cancer lesions.

Keywords: Breast imaging, Computed aided diagnosis, Medical image processing, Medical imaging.

1. Introduction

Based on the recent Global Cancer Statistics, in 2020, there were a total of 11.7% breast cancer cases and approximately 19.3 million were new cases [1]. Breast cancer is the most common cancer among women and is associated with a high death rate. [2, 3]. Early detection can provide effective patient management and treatment. Digitised mammograms are the most reliable and effective way of breast screening for early detection of breast cancer [4]. Digitised mammograms can also show different types of breast cancer lesions such as benign and malignant lesions [5]. Determining both lesions will help to provide the most appropriate treatment method and thus, this increases the survival rate [6]. However, digitised mammograms with low image quality will reduce lesion detectability and differentiation [7]. As a result, many patients will be asked to undergo further investigations such as an invasive biopsy. Therefore, a Computer-Aided Diagnosis (CAD) system has been developed for

* Corresponding author
E-mail address: harithsuradi@gmail.com

Manuscript History:

Received 15 June, 2021, Revised 16 August, 2021, Accepted 17 August, 2021, Published 31 October, 2021
Copyright © 2021 UNIMAS Publisher. This is an open access article under the CC BY-NC-SA 4.0 license.
<https://doi.org/10.33736/jaspe.3517.2021>

assisting radiologists to identify breast cancer lesions better at the early stage without undergoing invasive biopsies.

The CAD system has been developed to detect mass or microcalcifications clusters [8]. The most challenging problem of the CAD system is to detect the mass in low contrast digitised mammograms associated with complex shapes and undesired surrounding breast tissues. The major purpose of the CAD system is to reliably and accurately detect and identify the location of breast cancer lesions. In previous literature, the use of CAD systems had shown a good impact especially for early detection of breast cancer [9]. There are recent advancements in developing a CAD system for the detection of breast cancer in the last few decades [10, 11]. The development of the CAD system is mainly to improve breast cancer detection, particularly in the recognition of masses and calcifications. Nonetheless, certain areas of research in the CAD system of breast cancer still require further attention to improve accuracy. The major drawbacks for the CAD system of detection of breast cancer are poor image enhancement techniques for improving the images of digitised mammograms. For example, the CAD system still has limitations with low contrast and lack of preservation image details [9]. Thus, it will lead to double reading and increased time for detecting the location of breast cancer. A study by Makandar A, Halali B (2016) showed the CAD system for detection of breast cancer using median filter dan Contrast Limited Adaptive Histogram Equalization (CLAHE) of enhancement process and Otsu thresholding segmentation [12]. This system still has limitations to detect breast cancer due to poor intensity images of enhanced digitised mammograms.

In this paper, a CAD system with an automated classification of breast cancer lesions is proposed, consisting of low computational time and using rule-based classification with a newly proposed image enhancement technique for digitised mammograms. The proposed CAD system was found to be able to enhance the digitised mammograms by using Suradi SH, Abdullah KA, Isa NAM (2021) previous work of an improved image enhancement technique for digitised mammograms namely, Fuzzy Anisotropic Diffusion Histogram Equalization Contrast Adaptive Limited (FADHECAL) to reduce noise in digitised mammograms while preserving the contrast and the brightness. Breast cancer lesions segmentation using Multilevel Otsu Thresholding was also applied to detect and segment the mass on the digitised mammograms [13].

2. Materials and Methods

2.1. CAD system of Automated Classification of Breast Cancer Lesions

Four main blocks were used to automate the classification of breast cancer lesions; (i) Pre-processing and Enhancement block; (ii) Segmentation block; (iii) ROIs Extraction block; and (iv) Classification block. Figure 1 illustrates the proposed CAD system's main block diagram. First, the system read the digitised mammograms from the Mini-MIAS database. Then, these digitised mammograms underwent pre-processing and enhancement process by FADHECAL to improve the quality of images. The enhanced digitised mammograms were segmented using Multilevel Otsu Thresholding to segment the Region of Interests (ROIs) for breast cancer lesions. The final output of the segmentation process was then transferred to ROIs Extraction block to extract the types of breast cancer lesions based on specific features such as area, circularity and solidity. The ROIs were classified into two types of lesions - benign and malignant.

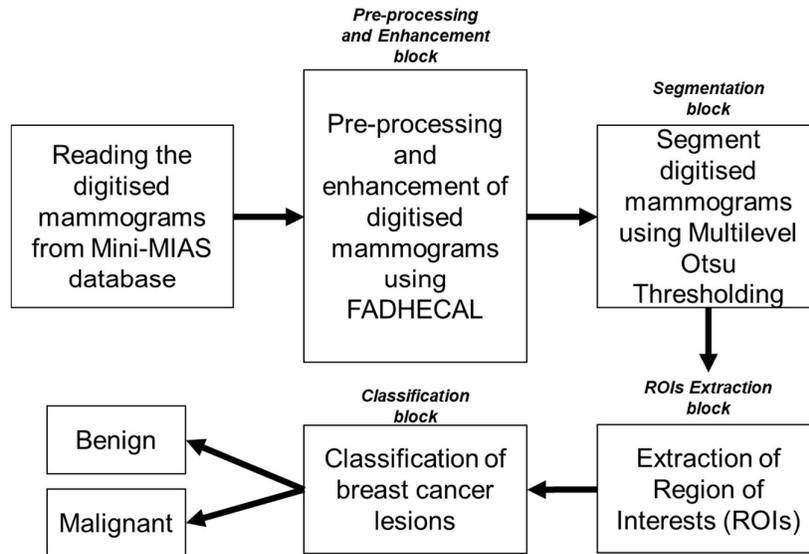


Figure 1. Overall block diagram for proposed CAD system of automated classification of breast cancer lesions.

2.1.1 Database of Digitised Mammograms

Digitised mammograms from the Mini Mammographic Image Analysis Society (MIAS) database were used in this study [14]. The database consisted of 322 digitised mammograms with normal, masses and microcalcifications cases. Only 30 digitised mammograms were included in this study due to related abnormalities within. The selected digitised mammograms consisted of three different background tissues such as fatty, fatty glandular and dense glandular tissues. Each digitised mammogram was transformed 8-bit grey levels quantisation with 256 different levels. The size for each digitised mammogram was 1024 X 1024 pixels. All digitised mammograms were in portable grey map (PGM) format. Figure 2 shows the digital mammograms from the Mini-MIAS database with different types of background tissues.

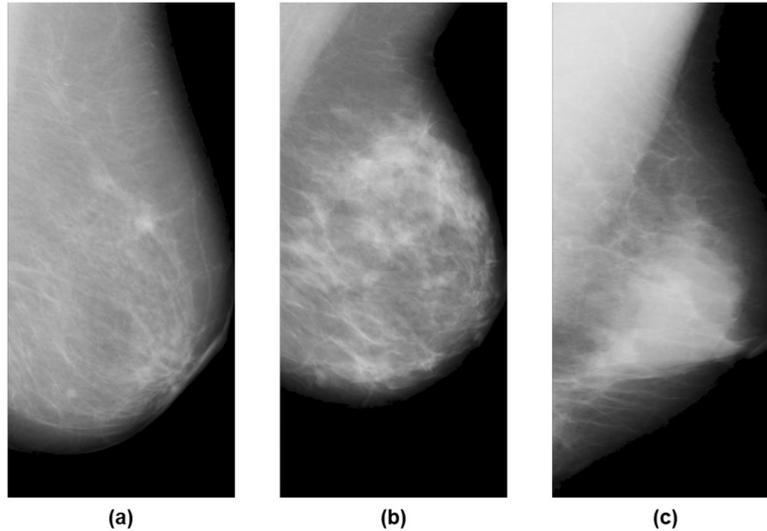


Figure 2. Digitised mammograms from Mini MIAS database with different background tissues; (a) Fatty tissue; (b) Fatty-glandular tissue and (c) Dense-glandular tissue.

2.1.2. Pre-processing and Enhancement using FADHECAL

The pre-processing of digitised mammograms was performed as the initial process to remove the unwanted region, such as the labels to improve the contrast appearances. FADHECAL is a hybrid enhancement technique that combines fuzzy-based, histogram-based enhancement techniques and an Anisotropic Diffusion Filter (ADF) is embedded to reduce the noise while preserving the image details. A Fuzzy Clipped Inference System (FCIS) was employed in this FADHECAL technique to automate the clip-limit selection for the enhancement process. The FCIS had been set up with two input measures, namely the contrast (C) and the entropy (E). The flowchart of FADHECAL enhancement for digitised mammograms is shown in Figure 3. The FADHECAL technique used Equation 1 to calculate the clip limit to enhance the digitised mammograms.

$$FADHECAL = \left\lfloor \frac{\varphi}{256} \right\rfloor + \left[FHCL \cdot \left(\varphi - \left\lfloor \frac{\varphi}{256} \right\rfloor \right) \right] \dots\dots\dots (1)$$

Where FHCL is for fuzzy histogram clip limit (ranges from 0 to 0.1), and $\lfloor . \rfloor$ stands for truncating value to the nearest integer. φ is the block size's product, the pixel intensity values vary from 0 to 255, as indicated by the value 256.

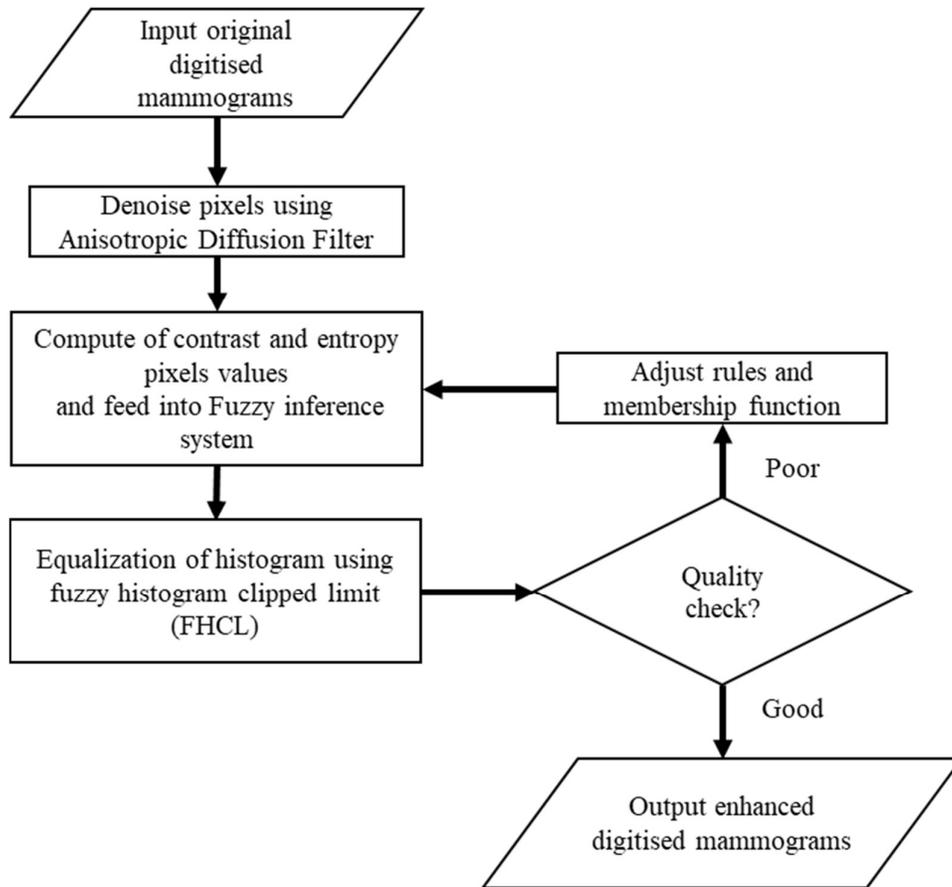


Figure 3. The flowchart of FADHECAL enhancement for digitised mammograms.

2.1.3. Segmentation using Multilevel Otsu Thresholding

A process to fraction or partition the image into regions is known as image segmentation. This technique was used to find and extract details in the image as Region of Interests (ROIs). The thresholding method was the simplest and one of the most widely used techniques for image segmentation due to its simplicity [15]. By assigning an intensity value to each pixel, thresholding was used to separate an ROI from its background, allowing each pixel to be categorised as an ROI or a background.

The segmentation process must satisfy certain conditions to obtain the region of breast masses. The selection region of interest for breast masses in digitised mammograms could identify the types of breast cancer lesions. Digitised mammograms often had abrupt changes in the intensity peak values [16]. Thus, it led to misclassifying the breast cancer lesions region due to different types of breast tissue densities. Multilevel Otsu Thresholding Segmentation could overcome this problem by segmenting the region of breast cancer lesions at different intensities levels. Multilevel Otsu thresholding uses an appropriate threshold (T), as in Equation 2.

$$y(x, y) = \begin{cases} 1, & \text{if } 130 \leq T \leq 211 \\ 0, & \text{otherwise} \end{cases} \dots\dots\dots (2)$$

$$y(x, y) = \begin{cases} 1, & \text{if } 130 = T = 211 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Where $y(x, y)$ is the binary image and $I(x, y)$ is the pixel intensity and (x, y) is the image location. If the image pixels have values larger or smaller than the set-up threshold values, the pixels show a white region or '1'. Otherwise, the pixels show a black region or '0'.

2.1.4. Region of Interests (ROIs) Extraction of Breast Cancer Lesions

Region of Interests (ROIs) extraction was used for digitised mammograms analysis. This method of morphological operations was used to extract the breast lesions within the ROIs of blob properties. In this study, this method was also used to help structure the elements and measure the shape of the image. It could refine the characteristics of the ROIs and assist in the removal of undesirable areas [17]. The parameters of the blobs segmented from digitised mammograms were computed using specific features of areas, circularity, and solidity as described:

i. Area of blob

The area was used to measure the number of pixels that were occupied by the breast cancer lesions. The white pixels represented the ROIs of breast cancer lesions.

ii. Circularity of blob

The circularity was used to measure the roundness, i.e. excluded the local irregularities. The circularity was obtained by calculating the ratio area of an object to the area of a circle with the same convex perimeter as shown in Equation 3.

$$\text{Circularity} = \frac{4\pi \cdot \text{Area}}{(\text{convex. perimeter})^2} \dots\dots\dots (3)$$

iii. Solidity of blob

The solidity was used to measure the density of ROIs for breast cancer lesions. The ratio of the ROI area to the area of the ROI convex yielded the solidity measurement. Equation 4 describes the solidarity calculation.

$$\text{Solidity} = \frac{\text{Area}}{\text{convex area}} \dots\dots\dots (4)$$

3. Results and Discussion

Medical image classification can be considered as the main technique of Computer-Aided Diagnosis (CAD) in the digital image processing system. In this study, CAD was used to aid the detection of breast cancer lesions by classifying the Region of Interests (ROIs) extraction known as a blob. This CAD was used on digitised mammograms to classify the blob as benign or malignant lesions. Table 1 shows the results of ROIs extraction from the selected digitised mammograms with different background tissues and degrees of severity. The classifications of the blob (i.e. benign or malignant) were performed based on three (3) conditions; (i) Area of the blob; (ii) Solidity of the blob; and (iii)

Circularity of blob. Figure 4 shows the classification criteria for the breast cancer lesions using rule-based.

Table 1. Results of ROIs extraction from the selected digitised mammograms

No.	Type of lesions	Background tissues	Area of blob	Circularity of blob	Solidity of blob	
1	Benign	Fatty	1,127	2.6	0.9	
2			5,272	3.0	0.9	
3			1,150	2.4	0.8	
4			1,974	2.6	0.8	
5			2,387	2.5	0.8	
6			2,762	2.9	0.8	
7			5,755	2.9	0.8	
8			Fatty-glandular	2,795	2.4	0.8
9			2,536	2.7	0.9	
10			3,868	3.0	0.9	
11			4,498	3.5	0.8	
12			Dense-glandular	1,129	3.3	0.9
13				7,281	3.1	0.9
14				4,345	3.2	0.9
15				4,657	3.4	0.9
16	Malignant	Fatty	3,977	4.1	0.8	
17			5,963	3.7	0.8	
18			5,884	4.1	0.7	
19			3,018	3.9	0.8	
20			2,876	4.2	0.8	
21			1,952	3.9	0.7	
22			4,304	4.3	0.8	
23			Fatty-glandular	2,752	4.1	0.7
24			3,197	3.8	0.8	
25			3,534	3.9	0.8	
26			1,496	4.5	0.7	
27			4,734	4.4	0.8	
28			Dense-glandular	3,814	3.9	0.7
29			2,577	4.2	0.8	
30			2,842	4.6	0.7	

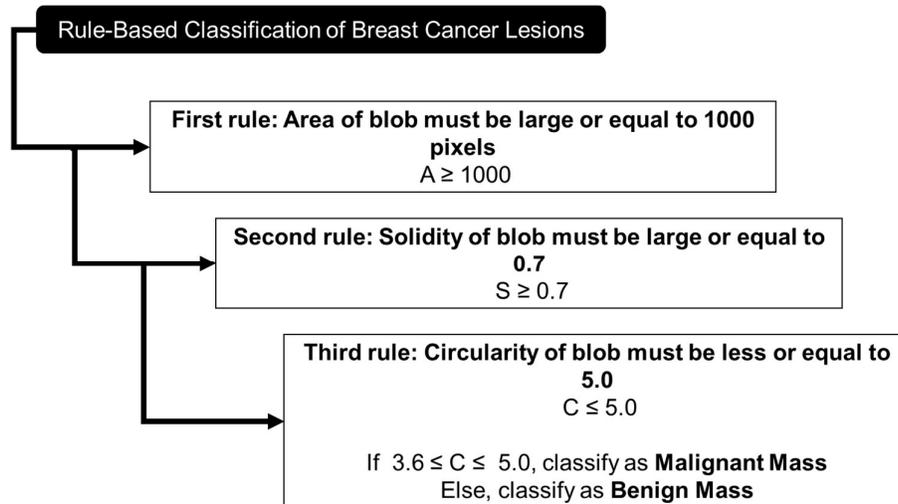


Figure 4. The classification criteria for breast cancer lesions using the rule-based method.

The area of the blob was used as the first rule to classify the breast cancer lesions. The blob area that was equal to or larger than 1000 pixels was classified as suspected of tumour or mass. The value of blob area was chosen based on digitised mammograms with abnormal conditions from the mini-MIAS database. As a result, the blob area value was chosen as the cut off point for excluding any segmented breast tissues that were unrelated to the tumour site. Blobs that met these criteria were assigned to a tumour or mass, and the remaining blobs were removed. Solidity was used as the second rule of blob criteria to classify breast cancer lesions. Solidity showed the fractions of a region as compared to its convex hull. The object shapes such as a solid circle or square were considered to have solidity. The values of solidity ranged from 0 to 1; whereby 1 was high solidity and 0 was not solid. Tumours had been considered as solid and they occupied most regions within the breast tissues. All abnormal conditions of digitised mammograms containing tumours or masses with an average solidity value greater than 0.7 were analysed in this study.

Circularity was used to measure the circular blob to classify the breast lesions. In the digitised mammograms, the blobs that contained tumours could appear in a well-defined or circular. In this study, most of the blobs with tumours or masses had circularity values of less or equal to 5.0. Therefore, the circularity value of 5.0 was chosen as the reference value. Another circularity value was also set as the reference value to distinguish between benign and malignant lesions. Blobs with a circularity value of less than 3.5 were classified as benign tumours, whereas those with a value of 3.6 to 5.0 were classified as malignant tumours. This condition range was established based on the fact that benign tumours had a rounder and more circular appearance than malignant tumours.

The classification was used to determine whether the area of blob or ROIs had a benign or malignant type of breast cancer lesions. If the area of blobs or ROIs had fulfilled the three specific criteria that were discussed above, tumours were suspected. Figures 5 and 6 show the digitised mammograms with benign and malignant masses respectively. Figures 5(a) and 6(a) show the original digitised mammograms with low contrast and suspicious mass region. After performing the FADHECAL technique as shown in Figures 5(b) and 6(b), it was found the mass had visualized and enhanced the details of suspicious mass. Thus, it increased the sensitivity of the subsequent segmentation process. The resultant of segmented images as shown in Figures 5(c) and 6(c) demonstrated the blob or ROI extraction that fulfilled all three criteria of classification of breast cancer lesions. Figure 5(c) shows the blob or ROI that fulfilled the rule or criteria of blob circularity less or equal to 3.5 (image = 3.0) and was classified as benign mass. Meanwhile, Figure 6(c) shows blob or

ROI that fulfilled the rule or criteria of blob circularity in a range from 3.6 to 5.0 (image = 4.2) and was classified as malignant mass. As shown in Figures 5(d) and 6(d), the location, size and shape of breast cancer lesion of mass regions were successfully detected and preserved.

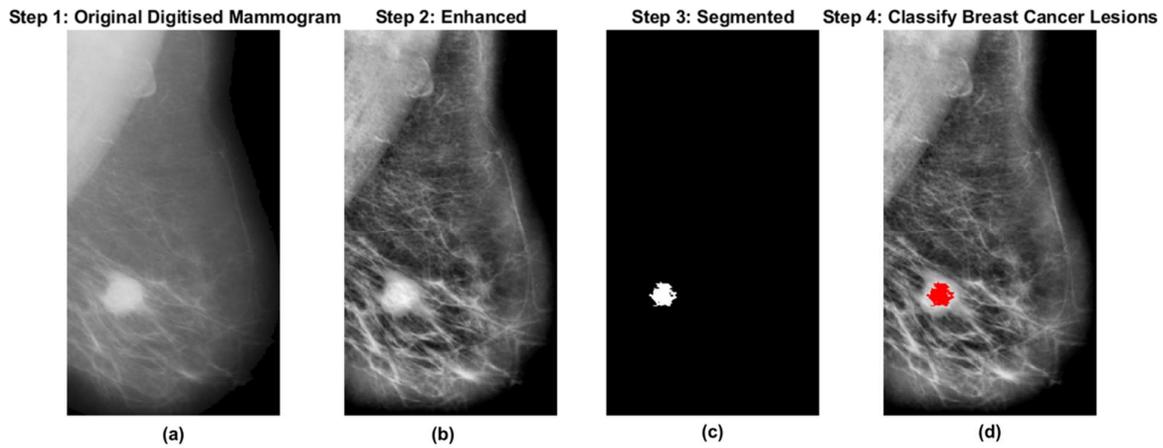


Figure 5. Results of detection of benign mass in a digitised mammogram; (a) Original digitised mammogram; (b) Enhanced digitised mammogram; (c) ROI extraction and (d) Location of benign mass in the digitised mammogram.

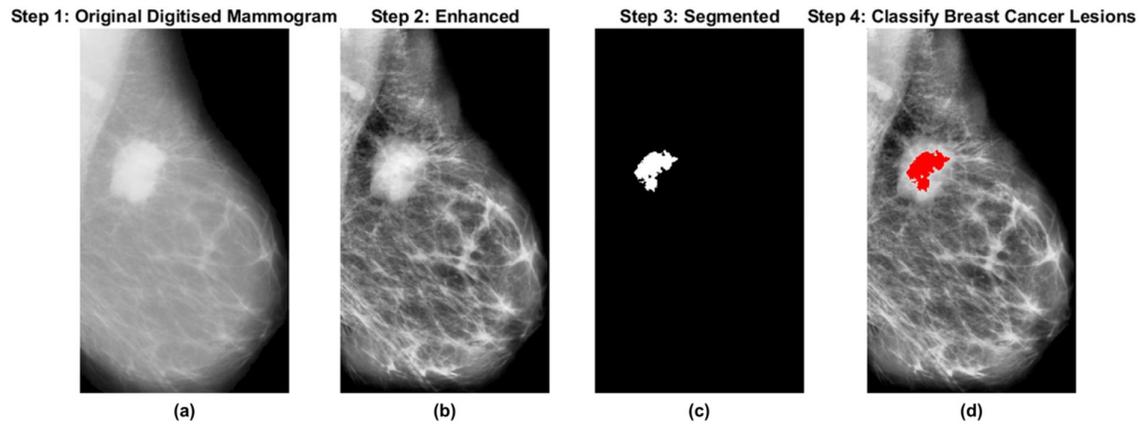


Figure 6. Results of detection of malignant mass in digitised mammogram; (a) Original digitised mammogram; (b) Enhanced digitised mammogram; (c) ROI extraction and (d) Location of malignant mass in the digitised mammogram.

After applying the proposed CAD system on all the selected digitised mammograms, quantitative analysis was carried out by employing the accuracy of the proposed algorithm as follows.

$$\text{Accuracy} = \frac{\text{Detected true positive of digitised mammograms}}{\text{Total of digitised mammograms}} \times 100\% \dots (5)$$

$$\text{Accuracy} = \frac{29}{30} \times 100\% = 96.67\%$$

The results showed the proposed CAD system had obtained a good performance for the detection of mass breast cancer lesions with 96.67% of accuracy. All the results were validated with the ground truth from the mini-MIAS database. From the 30 selected digitised mammograms, only one digitised mammogram was missed by the proposed CAD system due to the dense breast tissue surrounded by a mass of breast cancer lesions. For future research, the investigation needs to be done with a real case study (local database) with a higher number of digitised mammograms to improve the accuracy of the proposed CAD system.

4. Conclusion

Digitised mammograms can be difficult to interpret and diagnose due to low contrast appearances. Therefore, an automated CAD classification of breast cancer lesions has been proposed for digitised mammograms. This CAD system showed superior results that effectively detects and segments the masses of breast cancer lesions. The accuracy of CAD showed an excellent percentage of 96.67%.

References

- [1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3), 209-249. <https://doi.org/10.3322/caac.21660>
- [2] Suppaya, K., Nasir, F. M., & Ab Ghani, A. (2020). Variations of Bi-Rads 5 in Mammography by Age, Ethnicity, and Breast Density: A Retrospective Study in University Malaya Medical Centre. *Asian Journal of Medicine and Biomedicine*, 4(SI 1), 11-16. <https://doi.org/10.37231/ajmb.2020.4.SI%201.394>
- [3] Ravi, D.A. and Ismail, N. F. (2021). Knowledge And Awareness Of Breast Cancer And Mammography Among Women In Klang, Selangor. *Malaysian Journal of Applied Sciences*, 6(1), 15-20.. <https://doi.org/https://doi.org/10.37231/myjas.2021.6.1.265>Isa, N. A. M., & Siong, T. S. (2012). Automatic segmentation and detection of mass in digital mammograms. *Recent researches in communications, signals and information technology*, 143-146. ISBN: 978-1-61804-081-7
- [4] Helvie, M. A. (2010). Digital mammography imaging: breast tomosynthesis and advanced applications. *Radiologic Clinics*, 48(5), 917-929. DOI:<https://doi.org/10.1016/j.rcl.2010.06.009>
- [5] Paramkusham, S., Rao, K. M., & Rao, B. P. (2013, September). Early stage detection of breast cancer using novel image processing techniques, Matlab and Labview implementation. In *2013 15th International Conference on Advanced Computing Technologies (ICACT), IEEE*, 1-5. <https://doi.org/10.1109/ICACT.2013.6710511>
- [6] Abdallah, Y. M., Elgak, S., Zain, H., Rafiq, M., Ebaid, E. A., & Elnaema, A. A. (2018). Breast cancer detection using image enhancement and segmentation algorithms. *Biomedical Research*, 29(20), 3732-3736. <https://doi.org/10.4066/biomedicalresearch.29-18-1106>
- [7] ARazek, N. M. A., Yousef, W. A., & Mustafa, W. A. (2013). Microcalcification detection with and without CAD system (LIBCAD): A comparative study. *The Egyptian Journal of Radiology and Nuclear Medicine*, 44(2), 397-404. <https://doi.org/10.1016/j.ejrnm.2013.01.009>
- [8] Masud, R., Al-Rei, M., & Lokker, C. (2019). Computer-aided detection for breast cancer screening in clinical settings: scoping review. *JMIR medical informatics*, 7(3), e12660. doi: 10.2196/12660
- [9] Hadjiiski, L., Chan, H. P., Sahiner, B., Helvie, M. A., Roubidoux, M. A., Blane, C., ... & Shen, J. (2004). Improvement in radiologists' characterization of malignant and benign breast masses on serial mammograms with computer-aided diagnosis: an ROC study. *Radiology*, 233(1), 255-265. <https://doi.org/10.1148/radiol.2331030432>
- [10] Baker, J. A., Rosen, E. L., Lo, J. Y., Gimenez, E. I., Walsh, R., & Soo, M. S. (2003). Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting

- architectural distortion. *American Journal of Roentgenology*, 181(4), 1083-1088. <https://doi.org/10.2214/ajr.181.4.1811083>
- [11] Makandar, A., & Halalli, B. (2016). Threshold based segmentation technique for mass detection in mammography. *J Comput*, 11(6), 472-478. <https://doi.org/10.17706/jcp.11.6.463-4712>
- [12] Suradi, S. H., Abdullah, K. A., & Isa, N. A. M. (2021, April). Breast Lesions Detection Using FADHECAL and Multilevel Otsu Thresholding Segmentation in Digital Mammograms. In *International Conference on Medical and Biological Engineering*, Springer, Cham, 751-759. https://doi.org/10.1007/978-3-030-73909-6_85
- [13] Clark, A.F. (2012) The mini-MIAS database of mammograms. <http://peipa.essex.ac.uk/info/mias.html>
- [14] Goh, T. Y., Basah, S. N., Yazid, H., Safar, M. J. A., & Saad, F. S. A. (2018). Performance analysis of image thresholding: Otsu technique. *Measurement*, 114, 298-307. <https://doi.org/10.1016/j.measurement.2017.09.052>
- [15] Don, S., Choi, E., & Min, D. (2011, September). Breast mass segmentation in digital mammography using graph cuts. In *International Conference on Hybrid Information Technology*, Springer, Berlin, Heidelberg, 88-96. https://doi.org/10.1007/978-3-642-24106-2_12
- [16] Guzmán-Cabrera, R., Guzmán-Sepúlveda, J. R., Torres-Cisneros, M., May-Arrijoja, D. A., Ruiz-Pinales, J., Ibarra-Manzano, O. G., ... & Parada, A. G. (2013). Digital image processing technique for breast cancer detection. *International Journal of Thermophysics*, 34(8-9), 1519-1531. <https://doi.org/10.1007/s10765-012-1328-4>