

AUTOMATED DETECTION OF BREAST CANCER'S INDICATORS IN MAMMOGRAM VIA IMAGE PROCESSING TECHNIQUES

Abstract

The presence of abnormalities in mammographic image is an important step in the detection and diagnosis of breast cancer. The indicators of cancer in mammogram are in forms of calcification, mass and stellate lesion. This paper proposed a two-stage procedure for the detection of these cancer's indicators. In the first stage, original images obtained from Mammographic Image Analysis Society (mini MIAS) database were pre-processed and enhanced using top hat filtering method and the enhanced images were segmented using Otsu's method. The second stage dealt with the extraction of features from the images. Four features were extracted and selected using Gray Level Concurrence Matrix (GLCM). The features extracted as well as selected include energy, homogeneity, contrast, and correlation. Subtractive clustering and fuzzy logic techniques were employed for the cancer's indicators classification in the mammograms. The implementation of the image processing techniques was done with matrix laboratory. The result showed that the first image was affected by calcification, the second image was affected by mass and the third image was affected by stellate lesion. The method presented in this paper could assist the medical staff and improve the accuracy of breast cancer detection.

Keywords: *Mammogram, Breast Cancer, Image Processing, Fuzzy Logic, Subtractive Clustering*

1.0 Introduction

Breast cancer, a malignant tumour in the glandular tissue of the breast is one of the major causes of deaths among women. These tumours are referred to as carcinomas which occur when the processes that control the normal cell growth break down, thus enabling a single abnormal cell to multiply at a rapid rate, thereby destroying a proportion of the normal breast tissue over time [11].

Different image modalities are used in the detection and evaluation of breast abnormalities. The most common of these modalities is called the mammography. Mammography is a special x-ray technique that is used to visualize soft tissues of the

breast as a means of screening women with breast cancer. The result obtained from this technique is called the mammogram. Mammogram is capable of showing three indicators that can help to determine the likelihood of cancerous cells in the mammalian breasts. These indicators include; the mass, the calcifications and the stellate lesion.

In spite of the benefits of the mammogram in the detection of abnormal cells in the breast, its accuracy is not guaranteed [8]. For instance, scientists estimate that mammograms miss about 25 percent of breast cancer in women of 40 to 49 years old and about 10 percent of breast cancer in older women [5]. Some experts have also raised concerns about the ambiguity of mammography and the potential harm that a misleading result may cause.

Owing to these limitations and some other drawbacks of the mammography, there is a need for better and automated methods that facilitate the detection of this terminal disease. One of these methods could be achieved through different image processing techniques. Image processing is a rapid growing area in Computer Science. Its growth has been triggered by technological advances in digital imaging, computer processors and mass storage device. The primary purpose of image processing is to convert images to information. Digital image processing allows the use of much more complex algorithm. These algorithms include; the lowest level algorithms, the middle level algorithms and the highest level algorithms. Different tasks such as denoising, edge detection, segmentation, edge linking, feature extraction and selection and so on could be achieved through algorithms.

As a result of the availability of these algorithms, this paper proposed an image-based, efficient and automated method that can detect the likely factors that determine the presence of cancerous cells in the mammalian breasts.

2. Related Concepts

2.1 Breast Cancer

A cancer cell has characteristics that differentiate it from normal tissues cells with respect to the cell outline, shape, structure of the nucleus, and most importantly its ability to metastasize or infiltrate. When this occurs in the breast, it is referred to as breast cancer. Cancer is not curable, however early detection leads to a much higher chance of recovery and lowers the mortality rate from this disease.

The mammalian breasts are made of lobes and ducts. Each lobe is made up of lobules, and at the end of the lobules are tiny bulbs that can produce milk. The lobes, lobules, and bulbs are connected by tiny tubes called ducts. Each breast of a mammal has 15 to 20 lobes. Most breast cancers occur in either the lobules or the ducts in the breast. The breast also has lymph vessels which lead to the lymph node under the arms. The lymph nodes, also found in other parts of the body, make up the lymphatic system which plays the critical role of helping the body fight infection. Breast cancer often spreads to the lymph nodes under the arm and then through the lymphatic system [1].

Breast cancer, also known as carcinoma, is a malignant growth that begins in the tissues of the breast [3]. Breast cancer involves the uncontrolled growth of abnormal cells that have mutated from normal tissues. This growth prevents the normal functioning of vital organs or spread throughout the body damaging essential systems. However, the term benign refers to a condition, tumour or growth that is not cancerous. It does not spread to other parts of the body or invade and destroy nearby tissue. Benign tumours usually grow slowly. In general, benign tumour or condition is not harmful. However, this is not the case if a benign tumour is big enough, its size and weight can press on nearby blood vessels, nerves, organs or otherwise cause problems [3]. Carcinomas or breast cancer may also spread or metastasize to other parts of the body.

2.2 Stages of Breast Cancer

Breast cancer takes years to develop. It is commonly classified into four stages according to the size of tumours and degree of cancer spread from the breast to other parts of the body. These stages are highlighted below:

- The first stage is a pre-cancerous stage called ductal carcinoma in situ (DCIS) when a pre-cancerous lesion has not developed into a cancer tumor. In the first stage, a 0-2centimeter tumor forms without spreading outside the breast. If the cancer is detected in this stage the five-year survival rate is 96% [7].
- In the second stage, the cancerous cells form new malignant foci in positive lymph nodes or the tumour enlarges to 2-5 centimeter. In this stage, the survival rate drops to 73% [2].
- In the third stage, a tumour is larger than 5 centimeters with positive lymph nodes, or a tumour has skin and chest wall involvement. The surgical intervention for

this stage would be quite heavy; it may need partial or total breast removal and lymph nodes dissections.

- In the fourth stage, obvious metastases to other organs of the body, most often the bones, lungs, liver, or brain occur and the five-year survival rate drops to 20% [2].

2.3 Breast Cancer' s Indicators in Mammogram

Mammogram can show three factors that indicate the presence of cancerous cells in the breast. These include:

- **Mass:** A breast mass can be defined as a localized swelling, protuberance, or lump in the breast, which usually is described by its location, size, shape, margin characteristics, and any other associated findings such as distortion and X-ray attenuation.
- **Calcifications:** Calcifications are very small bits of calcium that appears within the soft tissue of the breast [8]. They usually appear as white dots on the mammogram. Calcifications are divided into two kinds. These are macro calcification and micro calcification. Macro calcification is coarse calcium deposit in the breasts. They are usually caused as a result of natural aging of the breasts or they can be as a result of previous injuries or inflammation in the breast. Macro calcifications are usually harmless and are not linked to breast cancer. They are benign and do not require any form of monitoring or treatment. Micro calcifications on the other hand in mammograms can be used as the primary indicator of breast cancer detection [4]. They may appear alone or in clusters. Its layout and shape can assist radiologists to determine the occurrence of cancer in the mammogram.
- **Stellate Lesion:** The detection of stellate lesions is very important in the characterization of breast cancer [3]. Stellate lesion often appears as a central mass surrounded by spicules radiating outwards. In general, it is difficult to detect because it has an irregular centre with ill-defined borders radiating spicules that may extend from several millimeters to centimeters in size. As a cancer cell proliferates, it shows up as a star-shaped with spiky lines radiating in all directions from a central region. A white star shape is a characteristic of a

malignant stellate lesion whereas the black star indicates a radial scar and post-traumatic fat necrosis.

3. Proposed Model's Approach

The images used for this paper were obtained from the mini MIAS (Mammographic Image Analysis Society) database. The methodology employed in this work is illustrated in Figure 1. The proposed model is a two-stage detection procedure. The purpose of the first stage is to determine the location and border of the suspected areas in the mammogram and to obtain an enhanced image. The image enhancement was achieved via top hat filtering method and also used to segment objects that differ in brightness from the surrounding background in images with uneven background intensity while the Otsu method was used to segment the enhanced borders from the background image. Morphological closing was applied to the image result of the Otsu method to eliminate noise.

The second stage involved the extraction of four features from the images using Gray Level Concurrence Matrix (GLCM). The features extracted and selected included energy, homogeneity, contrast, and correlation. The classification of the cancer type was done using subtractive clustering and fuzzy logic techniques. The image processing techniques were done using matrix laboratory.

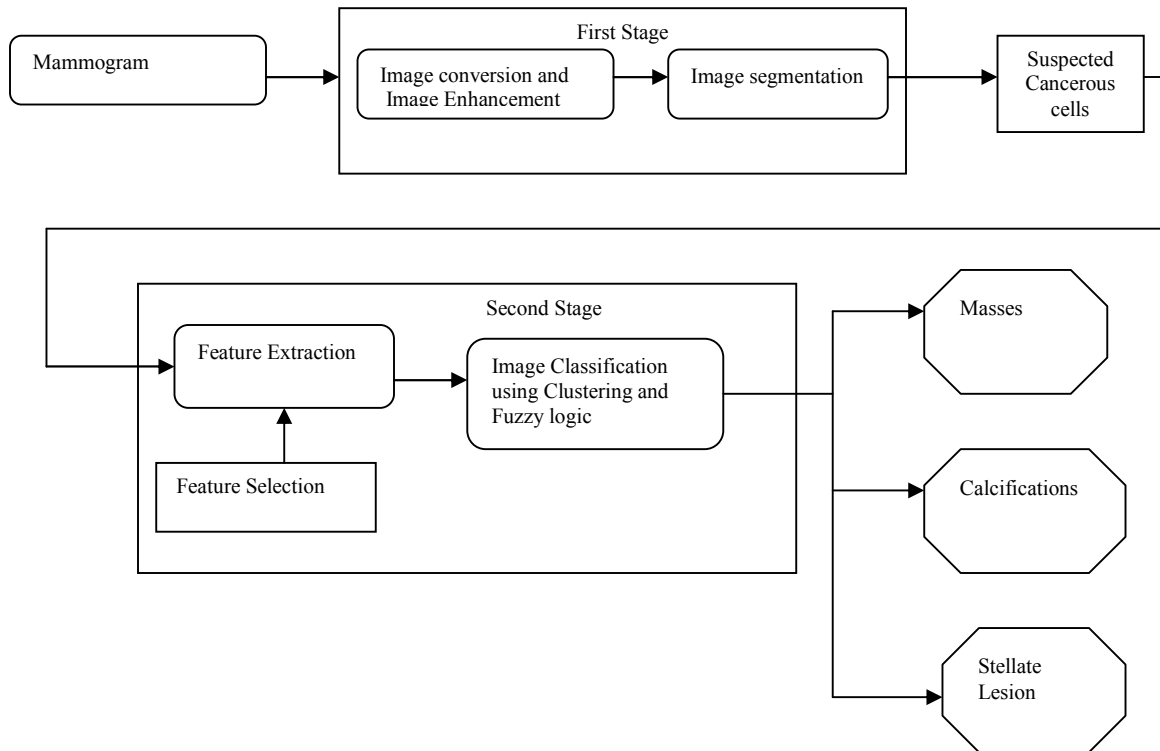


Figure 1: Proposed Model

3.1 Proposed Model's First Stage Procedure

3.1.1 Image Enhancement using Top Hat Filtering

The spot-like characteristics in the original mammogram images were enhanced using top hat filtering. The research employed the top hat filtering formula in Fu *et al.*, (2005) which is given as:

$$B = A - [(A \ominus SE) \oplus SE] \quad (1)$$

Where

A = input image (the X-ray mammogram). The image was loaded into the matlab environment using the 'imread' command.

B = processed image

SE = structure element. A structuring element is a matrix consisting of only 0's and 1's that can have any arbitrary shape and size. The pixels with values of 1 define the neighbourhood.

$\ominus$ = morphological erosion operation

$\oplus$ = morphological dilation operation

$-$ = image subtraction

3.1.2 Image Segmentation using Otsu's Method

The Otsu method was used to segment the enhanced borders from the background image. The Otsu's method was employed because it is the most accurate technique in detecting cancer in digital mammograms as it has shown a more satisfactory performance in the medical image segmentation (Mohammed, 2009). The equation for the Otsu method is given by:

$$d_w^2(T) = \omega_0 \sigma_0^2 + \omega_1 \sigma_1^2 \quad (2)$$

Where σ_0^2 and σ_1^2 are the variances of the pixels below and above the threshold T, respectively. ω_0 and ω_1 are defined as,

$$\omega_0 = \sum_{i=0}^{T-1} p(i) \quad (3)$$

$$\omega_1 = \sum_{i=T}^L p(i) \quad (4)$$

With $p(i)$ indicating the probabilities of the occurrence of gray level i in the image.

3.2 Proposed Model's Second Stage Procedure

3.2.1 Features Extraction and Selection

The second stage began with the extraction of four features from the images using Gray Level Concurrence Matrix (GLCM) and the equation is given as:

$$C(i, j) = |\{m, n\} : P(m, n) = i \text{ and } P(m + d \cdot \cos \theta, n + d \cdot \sin \theta) = j| \quad (5)$$

The co-occurrence matrix C , computed on a gray-level image P , is defined by a distance d and an angle θ . $C(i, j)$ is the number of times that the gray value i co-occurs with the gray value j in a particular spatial relationship defined by d and θ . The features extracted as well as selected included energy, homogeneity, contrast, and correlation. The goal of feature selection is to choose the optimal feature vector, consisting of the features that minimize the classification error as well as to differentiate among the different cancer's indicators. The model employed subtractive clustering algorithm that produced cluster centre values and sigma values which were used to perform the fuzzy logic rules for the classification of the cancer's indicators.

3.2.2 Image Classification using Fuzzy Logic

The clustering process produced cluster center values and cluster sigma which were used to perform the fuzzy logic rules. From these fuzzy rules, the memberships of each data on each cluster were performed, and the antecedents of each rule were quantified with union theory as follows:

(a) AND relation used min operation

(b) OR relation used max operation

This quantification process for each rule produced fitness-limit value of each rule. These values were weights for each fuzzy rule base to set the fuzzy output.

$$O = \frac{\sum_{r=1}^R (\alpha r * y_r)}{\sum_{r=1}^R \alpha r} \quad (6)$$

0 is the fuzzy output

α is the fitness-limit value

y_r is the output variable

R is the number of rules (equal to number of data cluster)

4. Results and Discussion

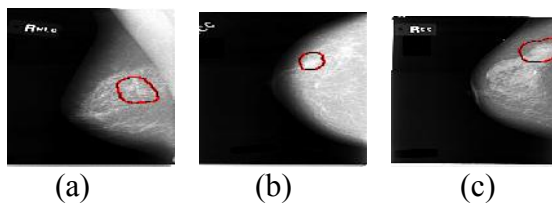


Figure 2: Original Images from mini MIAS database

The original images were read into matlab and then converted to gray images. This is because medical images are best worked upon when they are in gray scale. The gray scale equivalence are shown in figure 3(a, b & c).

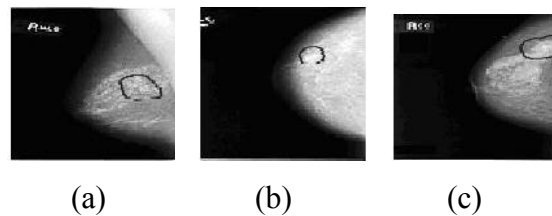


Figure 3: Gray scale images

The images were pre-processed using top hat filtering to remove noise and to get enhanced images. The result of the pre-processed operation is shown in figure 4(a, b & c)

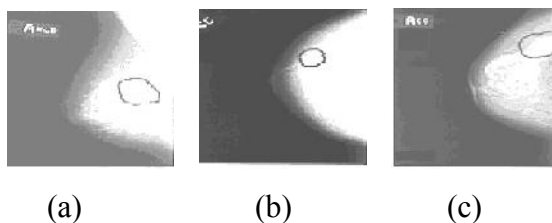


Figure 4: Result of the Image Enhancement

The Otsu's method was used to segment the region of interest (ROI) from the filtered images and the result is shown in figure 5(a, b & c)

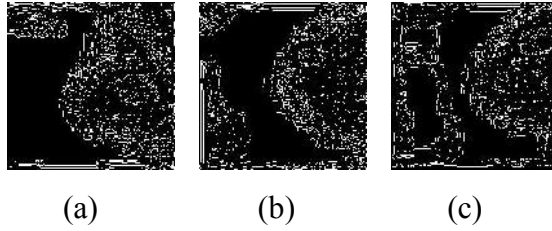


Figure 5: Result of the image segmentation

Four features were extracted from the images using Grey Level Concurrence Matrix (GLCM). The features extracted include contrast, homogeneity, correlation and energy and the result is shown in table 1.

Table 1: Result Of The Features Extracted

Images	Contrast	Homogeneity	Correlation	Energy
1	0.2520	0.6102	0.8632	0.6675
2	0.2171	0.6312	0.7015	0.5022
3	0.3418	0.9684	0.9253	0.5822

Contrast: This is a measure of intensity contrast between a pixel and its neighbour over the entire image.

Homogeneity: This measures the spatial closeness of the distribution of elements in the image to the diagonal.

Correlation: This is a measure of how correlated a pixel is to its neighbour over the entire image.

Energy: Energy is the sum of squared elements in GLCM. Energy may also be referred as uniformity or the angular second moment.

The features extracted from the segmented images were loaded into the matrix laboratory workspace as the input variables. The file was named datain. The cancer types identified from each of the mammograms derived from the MiniMias database that is; calcification, mass and stellate lesion were written into an output file. The relationship between the input variables and the output variable was modeled by first clustering the

data. Subtractive clustering was employed to identify natural groupings in data from the data set, thereby allowing concise representation of relationships embedded in the data.

Fuzzy logic was then employed to capture the broad categories identified during clustering into a Fuzzy Inference System (FIS). The FIS acts as a model that reflects the relationship between the input variables and output variables. The FIS is composed of inputs, outputs and rules. Each input in the FIS represents an input variable in the input dataset and each output in the FIS represents an output variable in the output dataset. Therefore each input and output was characterized by four membership functions and four rules were generated. The rules dictate the behaviour of the fuzzy system based on the inputs variables and output variable. The result of the FIS is shown in table 2.

Table 2: The Output Of The FIS

Image	Contrast	Homogeneity	Correlation	Energy	Output	Cancer Indicator
1	0.2520	0.6102	0.8632	0.6675	0.6002	Calcification
2	0.2171	0.6312	0.7015	0.5022	0.5130	Mass
3	0.3418	0.9684	0.9253	0.5822	0.7044	Stellate lesion

The result of table 2 showed that the FIS detected that image 1 is affected by calcification, image 2 is affected by mass and image 3 is affected by stellate lesion.

5. Conclusion

Breast cancer has been one of the major causes of death among women since the last decades and it has become an emergency for the healthcare systems of both developing and industrialized countries. Mammography is the main test used for screening and early diagnosis of breast cancer. However, the accuracy of mammography cannot be guaranteed. Consequently, this paper focused on the detection of breast cancer's indicators in mammography images using digital image processing techniques. The method presented in this paper will assist the medical staff in the early detection and evaluation of this terminal disease.

References

- [1] California Department of Public Health (2010). *Breast cancer*. California Department of Public Health, Cancer Surveillance Section, pp 115-118
- [2] Cardillo, F.A., Starita, A., Caramella, D. and Cilotti, A. (2001), "A neural tool for

- 307 breast cancer detection and classification in MRI,” 2001 Proceedings of the 23rd
 308 Annual EMBS International Conference, Oct 25-28, Istanbul, Turkey.
 309
- 310 [3] Chandra, J.S, Marcus, K., and Sivakumar, R. (2010), “Detection of Masses in Digital
 311 Mammograms”, International Journal of Computer and Network Security, Vol.
 312 2, No. 2, pp.78-86
 313
- 314 [4] Fu, J.C., Lee, S.K., Wong, S.T.C., Yeh J.Y., Wang, A.H., and Wu, H.K (2005),
 315 “Image Segmentation Feature Selection and Pattern Classification for
 316 Mammographic Microcalcifications”, Computerized Medical Imaging and
 317 Graphics Vol.29. pp. 419–429
 318
- 319 [5] Kerlikowske, K. (2008), Mammogram, Microsoft Encarta 2009. Redmond, WA:
 320 Microsoft Corporation.
 321
- 322 [6] Musoko, V. (2005), Biomedical Signal and Image Processing, A Ph.D Thesis.
 323 Chemical and process Engineering. Institute of Chemical Technology,
 324 Department of Computing and Control Engineering, Czech Republic, Prague
 325
- 326 [7] National Breast Cancer Foundation, U.S.A. [Online].
 327 Availble:http://www.nationalbreastcancer.org/early_detection/index.html
 328
- 329 [8] National Cancer Institute (2010), What You Need to Know About Breast Cancer, U.S.
 330 Department of Health and Human Services National Institute of Health,
 331 National Institute of Health, USA
 332
- 333 [9] Rafael, G. and Richard, W. (2004), An Introduction in Matlab: A Support Document
 334 For Digital Image Processing, Second Edition. Prentice Hall Saddle River, New
 335 Jersey 07458
 336
- 337 [10] Seamann, T. (2003), Digital Image Processing Using Local Segmentation, Ph.D
 338 Thesis, School of Computer Science and Software Engineering, Faculty of
 339 Information Technology Monash University , Australia
 340
- 341 [11] Spratt, K. (2008), Breast Cancer, Microsoft Encarta 2009 [DVD]. Redmond. WA:
 342 Microsoft Corporation
 343