

Automated Detection of Breast Cancer's Indicators in Mammogram via Image Processing Techniques

Olaleke J.O^{1*}, Adetunmbi A.O², Obe O.O², Iroju O.G¹

(Put * above the corresponding author and give telephone number, fax number and email ID in the footer)

¹Computer Science Department, Adeyemi College of Education, Ondo, Ondo State, Nigeria

²Computer Science Department, Federal University of Technology, Akure, Ondo State, Nigeria

ABSTRACT

ABSTRACT:

Aims: 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.

Methodology: Twenty images were used for the study. The images were obtained from Mammographic Image Analysis Society (mini MIAS) database. The images were pre-processed and enhanced using top hat filtering method and the enhanced images were segmented using Otsu's method. 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.

Result: The result showed that seven of the images were affected by stellate lesion, nine of the images were affected by microcalcification while four of the images were affected by mass.

Conclusion: 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. INTRODUCTION

Breast cancer, a malignant tumour in the glandular tissue of the breast is one of the major causes of deaths among women [1, 2,]. 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 [3]. Different image modalities are used in the detection and evaluation of breast abnormalities. The most common of these modalities is called the

* E-mail address: shollyjane@yahoo.com.

mammography. The result obtained from the mammography is called the mammogram. Mammogram is capable of showing indicators that help to determine the likelihood of cancerous cells in the breasts. These indicators include: the mass, the calcifications and the stellate lesion amongst others. 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 are very small bits of calcium that appears within the soft tissue of the breast [3]. 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. 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.

In spite of the benefits of the mammogram in the detection of abnormal cells in the breast, its accuracy is not guaranteed [2]. 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 [3]. In addition, mammograms contain low signal to noise ratio (low contrast) and a complicated structured background [3]. Owing to these limitations, there is a need for a better method that will facilitate the detection of this terminal disease. Hence, this paper proposes an image-based, efficient and automated method that will enhance the detection of cancerous cells in the breasts.

2. METHODOLOGY

Twenty mammographic images were used for this study. These images were obtained from the mini MIAS (Mammographic Image Analysis Society) database provided by the Mammographic Image Analysis Society (MIAS) in the United Kingdom (UK) which has 200 micron resolution[5]. Five of the images were affected by masses, seven images were affected by micro calcifications and eight of the images were affected by satellite lesions.

2.1 Description of the Tool Employed

The image processing tool box in matrix laboratory (matlab) was employed. This is because the Image Processing Toolbox provides a comprehensive set of functions for image manipulation, analysis, digital imaging, computer vision, and digital image processing. Its capabilities include colour space transformations, linear filtering, mathematical morphology, texture analysis, pattern recognition as well as image statistics.

2.2 Proposed Model

The proposed model is a two-stage detection procedure as shown in figure 1. 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 [5]. 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 [6]. The second stage involved the extraction of four features from the images using Gray Level Concurrence Matrix (GLCM) [7]. 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.

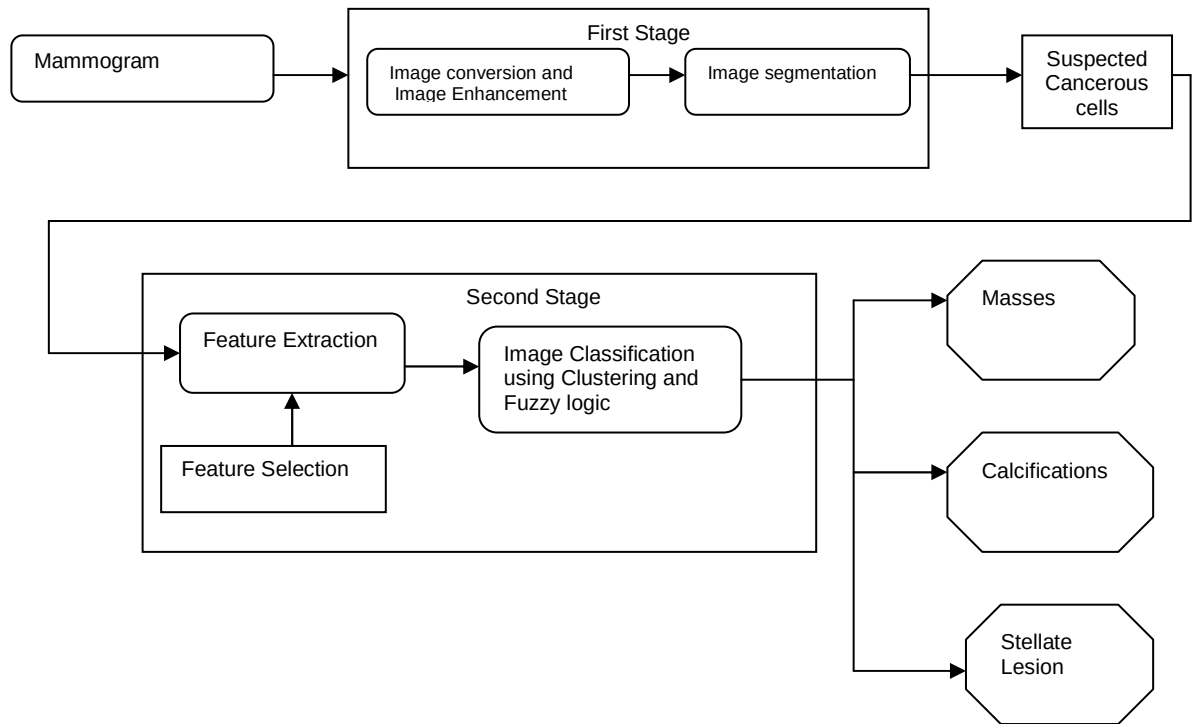
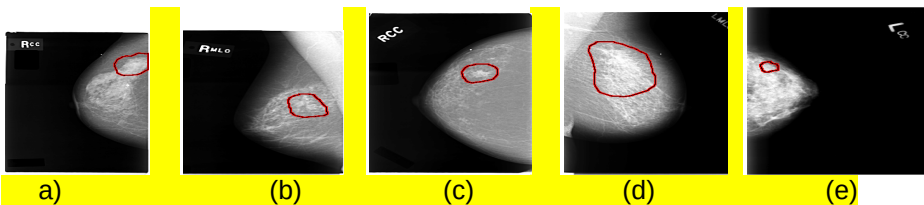


Fig. 1. Flow Chart of Proposed Model

3. RESULTS

Figure 2 shows the original images from miniMIAS database showing breasts affected by cancers.



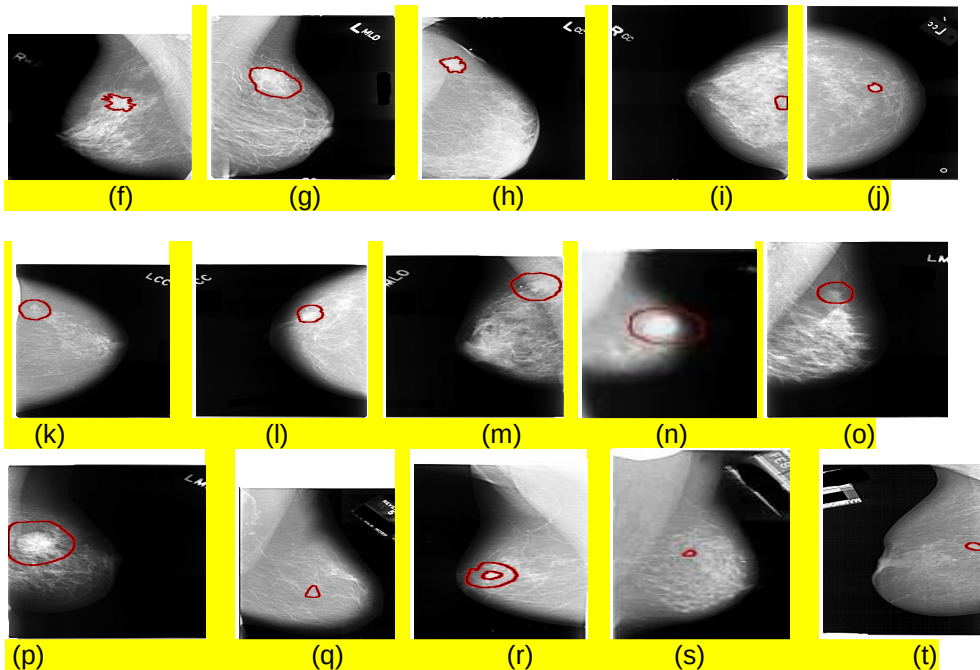


Fig.2. The original images from miniMIAS database

3.1 Image Selection Criteria

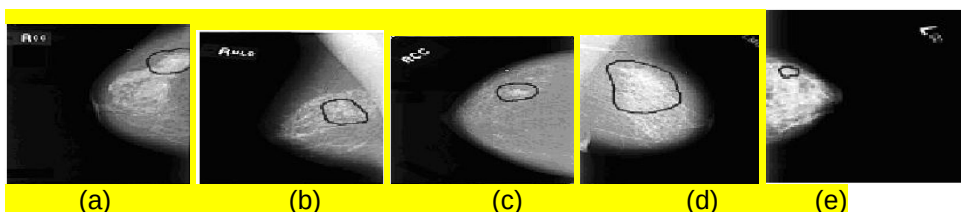
The original images were selected purposively from the miniMias database. This is to ensure that the selected images were affected by the three cancer indicators. Images a, c, f, h, k, p, r and t were affected by stellate lesion, images b, d, e, g, i, q and s were affected by calcification while j, l, m, n and p were affected by masses.

3.2 Image Conversion

The original images were converted to gray scale images. This is because medical images are better worked upon in gray scale and also to reduce the memory location of each image. The syntax used for the image conversion in matlab is shown in Equation (1) .

$$I = \text{rgb2gray}(\text{RGB}) \quad (1)$$

This converts the true colour image RGB (Red Green Blue) to the grayscale intensity image I. Rgb2gray converts RGB images to grayscale by eliminating the hue and saturation information while retaining the luminance [8]. For the conversion from RGB to grayscale image, the mathematical function is based on 30%Red + 59%Green +11%Blue. The result of the image conversion is shown in figure 3.



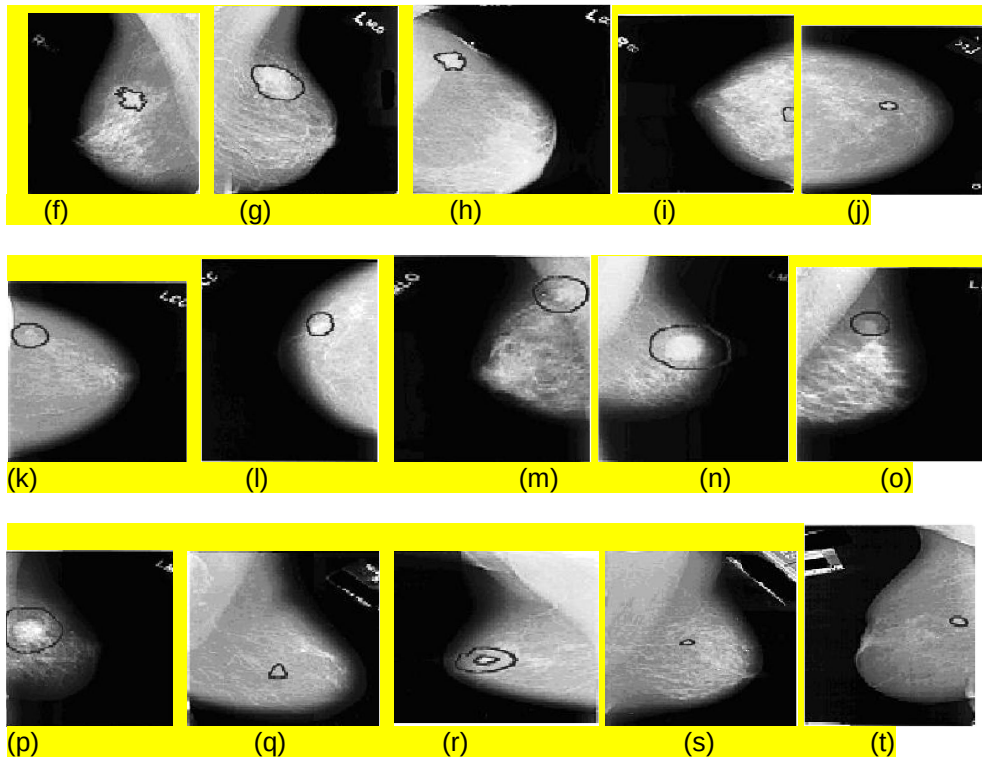


Fig.3. The Grayscale Images

3.3 Image Enhancement

The grayscale mammogram images were enhanced using top hat filtering [6]. The top hat filtering permits the extraction of light objects from an uneven background. The research employed the top hat filtering formula in [9] which is given as:

$$B = I - [(IOSE) \oplus SE]$$

Where

I is the gray scale image

B is the enhanced image

SE is the structuring element.

$\ominus$ is the morphological erosion operation which adds pixels to the boundaries of objects in the image [10]

$\oplus$ is the morphological dilation operation which removes pixels on the image boundaries

– is the image subtraction [10]

The syntax used for the image enhancement is shown in Equations (2) and (3) respectively.

$$se = strel('disk', 12) \quad (2)$$

$$J = imtophat(I, se) \quad (3)$$

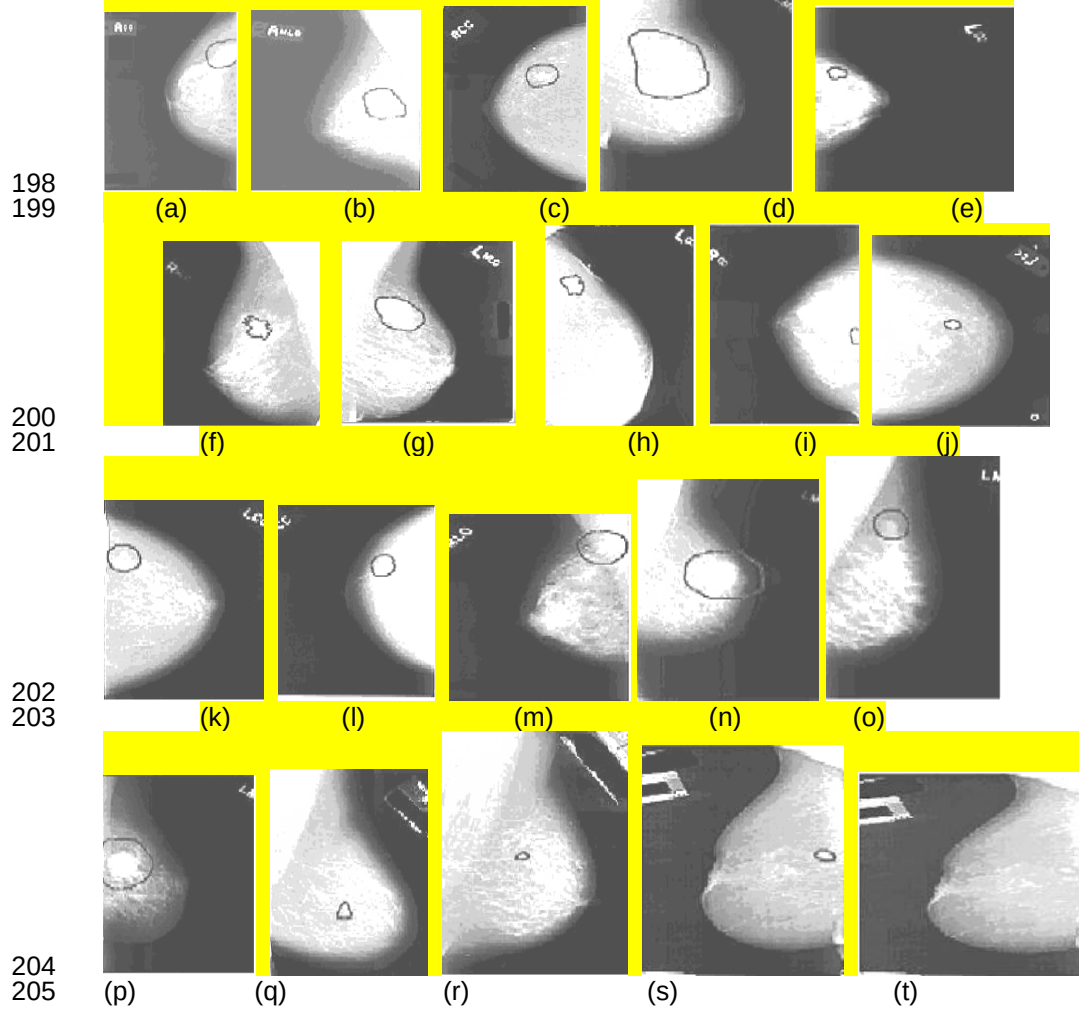
where

strel is a disk-shaped structuring element which removes the uneven background illumination from the images.

I is the grayscale image

196 J is the enhanced image.

197



202
203

204
205
206

207

Fig.3. The Enhanced images

208 3.4 Image Segmentation

209

210 The enhanced images were segmented using Otsu's technique [6]. This was used to
211 segment the enhanced borders from the background image. The Otsu's method was
212 employed because it is the most accurate technique in detecting cancer in digital
213 mammograms as it has shown a more satisfactory performance in the medical image
214 segmentation [10]. It has also been found to perform well compared to other thresholding
215 methods in segmenting the masses in digital mammogram [11]. The equations for the Otsu
216 method are shown in Equations (4), (5) and (6) [12].

$$217 \quad d_w^2(T) = \omega_0 \sigma_0^2 + \omega_1 \sigma_1^2 \quad (4)$$

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

220

$$\omega_0 = \sum_{i=0}^{T-1} p(i) \quad (5)$$

221

$$\omega_1 = \sum_{i=T}^L p(i) \quad (6)$$

222

223

224

225

226

227

228

229

230

231

232

233

234

235

With $p(i)$ indicating the probabilities of the occurrence of gray level i in the image. Sobel and Canny edge detection was then used to segment the enhanced borders from the background image. The Sobel edge detector applies Sobel approximation to the derivative of the image and detected edges whenever the gradient of reconstructed input image was at its maximum [5]. The Canny edge detector found edges by looking for local maximum of the gradient of unprocessed input image. In each edge detection algorithm, the gradient was calculated using the derivative of a Gaussian filter, and the output is a binary image, where 1 represents edges and 0 represents background [5]. The Sobel and Canny binary output images were post processed by the flood-filling operation to fill all objects with closed borders. The flood fill operation fills holes in the binary image. A hole is a set of background pixels that cannot be reached by filling in the background from the edge of the image [5]. They were then logically ORed together to produce a new image that includes false positives but discourages false negatives. This new image undergoes morphological closing to eliminate noise. The result of the image segmentation is shown in Fig.4.

236

237

238

239

240

241

242

243

244

245

246

247

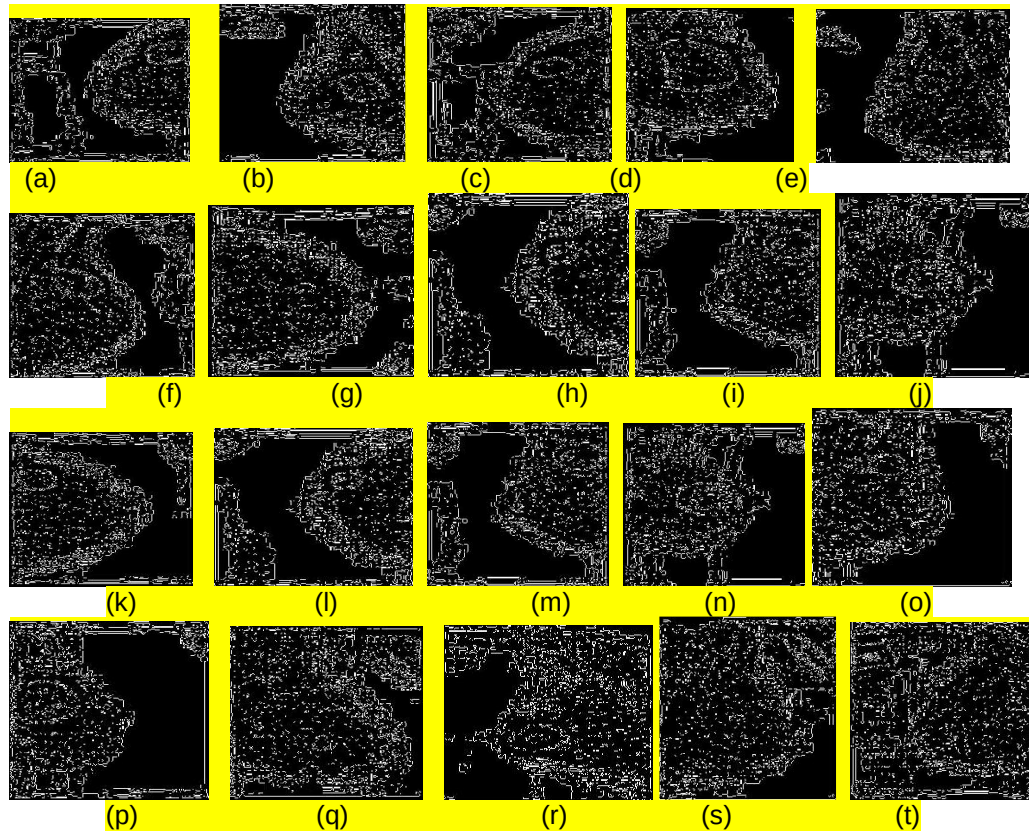


Fig. 4. The Segmented Images

3.5 Feature Extraction and Selection

Four features were extracted from the images using Gray Level Concurrence Matrix (GLCM). These features extracted include contrast which is a measure of intensity contrast between a pixel and its neighbour over the entire image, homogeneity which measures the spatial closeness of the distribution of elements in the image to the diagonal. The range of values is [0, 1], with the maximum being achieved when the image is a diagonal matrix. The homogeneity descriptor refers to the closeness of the distribution of elements in GLCM to the GLCM diagonal. Correlation which is a measure of how correlated a pixel is to its neighbour over the entire image and energy or uniformity which is the sum of squared elements in GLCM [9]. Energy may also be referred to as the angular second moment. 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 [9]. The equation for the features extraction is as shown in Equation (7) [12]

$$C(i, j) = |\{m, n\} : P(m, n) = i \text{ and } P(m + d \cdot \cos \theta, n + d \cdot \sin \theta) = j| \quad (7)$$

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 result obtained from the feature extraction is shown in Table 1.

A matrix laboratory file containing the features extracted that is contrast, homogeneity, correlation, and energy from the segmented images were loaded into the matrix laboratory workspace as the input variables. The file was named datain. The cancer indicators identified from each of the mammograms derived from the MiniMias database that is micro calcification, masses and stellate were written into an output file. The output file tagged dataout was loaded into the matlab workspace as the output variables. The two variables that is the input and the output variables loaded in the workspace has four columns representing the four input variables and one column representing the output variable respectively. The twenty rows in datain and dataout represent the number of observations or samples or data points available. A row in datain that is the input variables constitutes a set of observed values of the four input variables (contrast, homogeneity, correlation, and energy) and the corresponding row in dataout represents the value of cancer indicator given the observations made for the input variables.

The relationship between the input variables and the output variable was modeled by first clustering the data.

Table 1. Features Extracted

Images	Contrast	Homogeneity	Correlation	Energy
a	0.3418	0.9684	0.9253	0.5822
b	0.2520	0.6102	0.8632	0.6675
c	0.2741	0.6314	0.8710	0.6645
d	0.2513	0.6500	0.8512	0.6551
e	0.2704	0.6301	0.8711	0.6345
f	0.3015	0.9854	0.9668	0.5517
g	0.2815	0.6048	0.8519	0.6748
h	0.3398	0.9934	0.9550	0.5238
i	0.2679	0.6456	0.8602	0.6528
j	0.2553	0.6080	0.8718	0.6900
k	0.3458	0.9545	0.9420	0.5904
l	0.2171	0.6312	0.7015	0.5022
m	0.2302	0.6059	0.7204	0.5210
n	0.2050	0.6218	0.7311	0.5008
o	0.3015	0.9917	0.9358	0.5818
p	0.2213	0.6034	0.7032	0.5153
q	0.2917	0.6114	0.8500	0.6512
r	0.3405	0.9535	0.9732	0.5712
s	0.2708	0.6708	0.8614	0.6740
t	0.3005	0.9559	0.9645	0.5825

3.6 Image classification using subtractive clustering and fuzzy logic

Subtractive clustering was employed to identify natural groupings in data from the data set, thereby allowing concise representation of relationships embedded in the data [13]. Subtractive clustering is a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a dataset [13]. The syntax used for generating this is:

$[C, S] = \text{subclust}([\text{datain dataout}], 0.5)$

where C is the cluster center and S is the sigma value that specifies the range of influence of a cluster center in each of the data dimensions. All cluster centers share the same set of sigma values and 0.5 is the radius which marks the cluster's radius of influence in the input space. The result of the cluster center and sigma value from the subtractive clustering approach is as shown in Tables 2 and 3 respectively. S has 5 columns representing the influence of the cluster centers on each of the 5 dimensions. This clustering technique allowed the groupings of the input variables into broad categories of the output variable, that is, cancer indicators hence allowing for easy classification.

Table 2: Result of the Cluster Centers

Contrast	Homogeneity	Correlation	Energy	Output
0.2741	0.6314	0.8710	0.6645	0.6103
0.2213	0.6034	0.7032	0.5153	0.5108
0.3418	0.9684	0.9253	0.5822	0.7044
0.3015	0.9854	0.9668	0.5517	0.7014

309 Table 3 Result of the Sigma Value

310	Sigma Values
311	
312	0.0249
313	0.0689
314	0.0480
315	0.0334
316	0.0351
317	
318	
319	
320	
321	

322 Fuzzy logic was then employed to capture the broad categories identified during clustering
 323 into a Fuzzy Inference System (FIS). Genfis2 is the function that creates a FIS using
 324 subtractive clustering. Genfis2 employs subclust behind the scenes to cluster the data and
 325 uses the cluster centers and their range of influences to build a FIS. The FIS acts as a model
 326 that reflects the relationship between the input variables and output variables. The FIS is
 327 composed of inputs, outputs and rules. Each input in the FIS represents an input variable in
 328 the input dataset and each output in the FIS represents an output variable in the output
 329 dataset. Therefore each input and output was characterized by eight membership functions.
 330 The membership function type is Gaussian type membership function and the parameters of
 331 the membership function are [0.0051 0.2196], where 0.0051 represents the spread
 332 coefficient of the Gaussian curve and 0.2196 represents the center of the Gaussian curve.
 333 in1cluster1, in1cluster2, in1cluster3, in1cluster4, in1cluster5, in1cluster6, in1cluster7 and
 334 in1cluster8 respectively capture the position and influence of the clusters for the input
 335 variable contrast. Similarly, the other seven inputs follow the exact pattern mimicking the
 336 position and influence of the other seven clusters along their respective dimensions in the
 337 dataset. The rules dictate the behaviour of the fuzzy system based on the inputs variables
 338 and output variable [13]. Therefore, the number of rules equals the number of clusters and
 339 since eight clusters were created, eight rules were generated as shown in Table 4.

340
 341
 342 Table 4. The Rules Generated from the FIS
 343

Rules	Weights
If (Contrast is in1cluster1) and (Homogeneity is in2cluster1) and (Correlation is in3cluster1) and (Energy is in4cluster1) then (Cancer Type is out1cluster1)	1
If (Contrast is in1cluster2) and (Homogeneity is in2cluster2) and (Correlation is in3cluster2) and (Energy is in4cluster2) then (Cancer Type is out1cluster2)	1
If (Contrast is in1cluster3) and (Homogeneity is in2cluster3) and (Correlation is in3cluster3) and (Energy is in4cluster3) then (Cancer Type is out1cluster3)	1
If (Contrast is in1cluster4) and (Homogeneity is in2cluster4) and (Correlation is in3cluster4) and (Energy is in4cluster4) then (Cancer Type is out1cluster4)	1
If (Contrast is in1cluster5) and (Homogeneity is in2cluster5)	1

and (Correlation is in3cluster5) and (Energy is in4cluster5)	
then (Cancer Type is out1cluster5)	
If (Contrast is in1cluster6) and (Homogeneity is in2cluster6)	1
and (Correlation is in3cluster6) and (Energy is in4cluster6)	
then (Cancer Type is out1cluster6)	
If (Contrast is in1cluster7) and (Homogeneity is in2cluster7)	1
and (Correlation is in3cluster7) and (Energy is in4cluster7)	
then (Cancer Type is out1cluster7)	
If (Contrast is in1cluster8) and (Homogeneity is in2cluster8)	1
and (Correlation is in3cluster8) and (Energy is in4cluster8)	
then (Cancer Type is out1cluster8)	

344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363

The first rule indicates that if the inputs to the FIS, that is, contrast, homogeneity, correlation, and energy strongly belong to their respective cluster1 membership functions then the output that is cancer indicator must strongly belong to its cluster1 membership function. The (1) at the end of the rule is to indicate that the rule has a weight or an importance of 1. Weights take its values between 0 and 1. Rules with lesser weights will count for less in the final output. The significance of the rule is that it concisely maps cluster 1 in the input space to cluster 1 in the output space. Similarly, the other seven rules map cluster2, cluster3, cluster4, cluster5, cluster6, cluster7 and cluster8 respectively in the input space to cluster2, cluster3, cluster4, cluster5, cluster6, cluster7 and cluster8 respectively in the output space. If a data point closer to the first cluster has a strong membership to the first cluster, it is fed as an input to the FIS then rule1 will fire with more firing strength than the other seven rules. Similarly, an input with strong membership to the second cluster will fire the second rule with more firing strength than the other seven rules and so on. The output of the rules (firing strengths) is then used to generate the output of the FIS as shown in Table 5.

Table 5: The output of the rules (firing strengths) is then used to generate the output of the FIS

Images	Contrast	Homogeneity	Correlation	Energy	Output
b	0.2520	0.6102	0.8632	0.6675	0.6002
c	0.2741	0.6314	0.8710	0.6645	0.6103
d	0.2513	0.6500	0.8512	0.6551	0.6019
e	0.2704	0.6301	0.8711	0.6345	0.6015
f	0.3015	0.9854	0.9668	0.5517	0.7014
g	0.2815	0.6048	0.8519	0.6748	0.6040
h	0.3398	0.9550	0.9550	0.5238	0.7030
i	0.2679	0.6456	0.8602	0.6528	0.6066
j	0.2553	0.6080	0.8718	0.6900	0.6062
k	0.3458	0.9545	0.9420	0.5904	0.7080
l	0.2171	0.6312	0.7015	0.5022	0.5130
m	0.2302	0.6059	0.7204	0.5210	0.5268
n	0.2050	0.6218	0.7311	0.5008	0.5146
o	0.3015	0.9917	0.9358	0.5818	0.7027
p	0.2213	0.6034	0.7032	0.5153	0.5108
q	0.2917	0.6114	0.8500	0.6512	0.6010
r	0.3405	0.9535	0.9732	0.5712	0.7096
s	0.2708	0.6708	0.8614	0.6740	0.6192
t	0.3005	0.9559	0.9645	0.5825	0.6192

364 The result of Table 6 showed that the proposed system detected seven mammograms that
365 were affected by stellate lesion, nine mammograms affected by mircocalcification, and four
366 mammogram images were affected by mass. This is in contrast with the result of the
367 mammogram images in the MiniMias database which indicated that there were five
368 mammogram images affected by masses and seven affected by micro calcification and eight
369 were affected.

370 The result generated from FIS was also supported by the mathematical parameters
371 proposed by [7]. The experimental result was generated based on the following
372 mathematical parameters:

373
374
375
$$\text{Mass}(c, h, c_0, e) = \begin{cases} \text{if } 0.20 \leq c \leq 0.25 \\ 0.60 \leq h \leq 0.65 \\ 0.70 \leq c_0 \leq 0.85 \\ 0.50 \leq e \leq 0.60 \end{cases}$$

376
377
378
379
380
381
$$\text{calcification}(c, h, c_0, e) = \begin{cases} \text{if } 0.25 \leq c \leq 0.30 \\ 0.60 \leq h \leq 0.65 \\ 0.85 \leq c_0 \leq 0.90 \\ 0.65 \leq e \leq 0.70 \end{cases}$$

382
383
384
385
386
387
388
$$\text{stellate}(c, h, c_0, e) = \begin{cases} \text{if } 0.30 \leq c \leq 0.35 \\ 0.60 \leq h \leq 0.65 \\ 0.90 \leq c_0 \leq 1.00 \\ 0.50 \leq e \leq 0.60 \end{cases}$$

389
390
391
392
393
394 where c = contrast

395 h = homogeneity
396 c₀ = correlation
397 e = energy
398
399

400 4. EVALUATION

401
402 This section compares the new model with the existing one designed by [9]. The parameter
403 used for the comparative analysis of both systems was detection rate. The model developed
404 in the previous research [9] was only able to detect micro calcification in mammogram
405 images, while the proposed system is able to detect microcalcification, masses and stellate
406 lesions.

407 The performance of the systems detection rate was evaluated using two parameters, True
408 Positive (TP), False Positive (FP). A TP is obtained when a mammogram is correctly
409 detected as mass, stellate lesion and micro calcification as obtained in the MiniMias
410 database. When a normal mammogram is incorrectly classified, then it is defined as a FP.
411 The measures are based on the formula:

412
413

$$Detection\ Rate = \frac{TP}{TP + FP} \times 100\%$$

414

415 The percentage of detection rate of the FIS system is as computed below:

416 TP=18

417 FP=2

418 Hence, the detection rate of the FIS is 90%

419

420 Also, the percentage of detection rate of the existing system developed by [9] was
421 computed. The authors claimed that there were a total of 7531dataset, out of these, 1117
422 true micro calcification was detected. Hence, the percentage detection rate of the existing
423 system was 14.83%

424

425 The detection rate of the FIS is 90% while that of the existing system is 14.83%. This shows
426 that the detection rate of the system of the FIS system is greater than that of the existing
427 system. Hence, the performance in terms of cancer detection of the FIS system is higher
428 than that of the existing system.

429

430 5.0 DISCUSSION

431

432 A breast cancer detection system has been introduced in this paper. The breast cancer
433 detection system employed image processing techniques such as top hat filtering, Otsu
434 method, Canny and Sobel method, Gray Level Concurrence Matrix, subtractive clustering
435 and fuzzy logic techniques. The proposed breast cancer detection system is a two stage
436 detection procedure which was based on [9]. The major distinction between the existing
437 system and the proposed system is that the existing system was used for microcalcification
438 detection while the proposed system dealt with the detection of three cancer indicators which
439 are microcalcification, stellate lesion and masses. The proposed system was tested with
440 twenty mammographic dataset obtained from the miniMias database. These mammographic
441 images were affected by the three cancer indicators. Seven of the images were affected by
442 stallete lesion, seven of them were affected by microcalcifications while five of the images
443 were affected by mass. However, the result of the proposed system showed that seven of
444 the images were affected by stellate lesion; nine of them were affected by microcalcifications
445 while four of the images were affected by mass. This showed a discrepancy between the

446 result of original mammographic images and those of the proposed system. For instance,
 447 the result from the miniMias database showed that images c and j were affected by stellate
 448 lesion and mass respectively while the proposed system showed that images c and j were
 449 affected by microcalcification. The discrepancy is as shown in Table 6. In addition, the
 450 performance evaluation of the system was performed. The parameter used for the
 451 performance evaluation was detection rate. The result of the performance evaluation
 452 showed that the detection of the proposed system is 90%. However, the detection rate of the
 453 existing system is 14.83%. The high discrepancy in the detection date could be as a result of
 454 the large dataset used to test the existing system.

455
 456
 457

Table 6: Comparison from the MiniMias Database and the Proposed System

Images	Result from MiniMias database	Result of Proposed System
c	stellate lesion	microcalcification
j	Mass	microcalcification

6. CLINICAL APPLICATIONS OF THE PROPOSED SYSTEM

The Proposed system if adopted would aid medical practitioners in the following areas:

6.1 Facilitates the extraction of clinically useful information

During the process of diagnosis and treatment, patients routinely undergo some procedures which require diverse techniques. Clinical information is usually generated during these procedures. However, to make sense of this information, it is important to provide reliable computational tools that can automatically analyze and extract information from images. Hence, this system will aid clinicians extract useful information for clinical decision making.

6.2 Enhancement of Mammographic Images

Mammograms usually contain low signal to noise ratio (low contrast) and a complicated structured background. Hence, there is a need to enhance mammographic imaging. Thus, this system will aid in the enhancement of mammographic images by making it clearer. Thereby, increasing the quality of the images.

6.3 Interpretation of Mammographic Images

This system will help humans and computers to easily interpret mammographic images once the images have been enhanced.

7 LIMITATIONS OF THE PROPOSED SYSTEM

The proposed system extracted only four features namely contrast, homogeneity, correlation and energy from the mammographic images. However, images are composed of diverse features such as entropy, auto-correlation, sum of squares and variance, standard deviation, skewness, average and dissimilarity. The purpose of extracting these features is to minimize the classification error. Hence, the more features extracted, the less the classification error. This study is also limited to the detection of three cancer indicators which are mass, microcalcification and stellate lesion.

8 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 breast cancer detection system employed image processing techniques such as top hat filtering, Otsu method, Canny and Sobel detection, subtractive clustering and fuzzy logic. The detection system is a two stage detection procedure. The first stage determines the location and border of the suspected areas in the mammogram while the second stage involved the extraction of features from the images and the classification of the images into the three cancer indicators. Hence, this method can be used to assist radiologists by helping them to improve the accuracy of breast cancer diagnosis.

COMPETING INTERESTS

"Authors have declared that no competing interests exist."

AUTHORS' CONTRIBUTIONS

'OLALEKE J.O' initiated the study and wrote the first draft of the manuscript. Adetunmbi, A.O and Obe O.O supervised the study while Iroju O.G assisted in the design of the system.

REFERENCES

1. California Department of Public Health. Breast cancer. California Department of Public Health, Cancer Surveillance Section. 2010.
2. Radiological Society of North America. Mammography. Accessed June 4 2013. Accessed 29 March 2012. Available: <http://www.radiologyInfo.org>.
3. Saslow D, Boetes C, Burke W. American Cancer Society Guidelines For Breast Screening With MRI As An Adjunct Tomammography. A Cancer Journal for Clinicians. 2007;57:. 79-89
4. Mohiy H, Mohamed A, Walid D. Detection of Breast Cancer Tumor Algorithm using Mathematical Morphology and Wavelet Analysis. GVIP Conference, Cairo, Egypt. 2005.
5. Fei G. An Efficient Approach to Automated Segmentation in Medical Image Analysis. Master of Science in Electrical Engineering, Texas Technical University, USA.
6. Sheshadri, H.S. And Kandaswamy, A. Detection of Breast Cancer by Mammogram Image Segmentation. Department of Electronics & Communication Engineering, PSG College of Technology, Coimbatore-641004 Tamilnadu, India. 2005.
7. Khuzi AM, Besar R, Wan Z, Ahmad NN. Identification of Masses in Digital Mammogram Using Grey Level Co-Occurrence Matrices. Biomedical Imaging and Intervention Journal. 2009: 5(3).
8. Seamann, T. Digital Image Processing Using Local Segmentation. Ph.D Thesis, School of Computer Science and Software Engineering, Faculty of Information Technology Monash University , Australia. 2003
9. Fu JC, Lee SK, Wong STC, Yeh JY, Wang AH, Wu H.K. Image Segmentation Feature Selection and Pattern Classification for Mammographic Micro calcifications. Computerized Medical Imaging and Graphics. 2005; 29:419–429.
10. Duarte MA, Alvarenga AV, Azevedo, CM, Calas MJG., Infantosi, AFC, Pereira WCA. Segmenting mammographic microcalcifications using a semi-automatic procedure based on Otsu's method and morphological filters. Brazilian Journal of Biomedical Engineering. 2013;29(4), 377-388.
11. Mohammed WA, Kadah YM. Computer Aided Diagnosis of Digital Images. International Conference on Computer Engineering & Systems. 2009; 299-303.
12. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics. 1979;9(1):62-66.
13. Jain Anil K. Data Clustering: 50 Years Beyond K-Means. Department of Computer Science and Engineering, Michigan State University, USA. 2008.