Shape and Texture Features for the Identification of Breast Cancer

Abdulkader Helwan, *Member, IAENG*, Rahib Abiyev

Abstract— this paper aims to develop intelligent breast cancer identification system based image processing techniques and neural network classifier. Recently, many researchers have developed image classification systems for classifying breast tumors using different image processing and classification techniques. The challenge is the extraction of the real features that distinguish the benign and malignant tumor. The classification of breast cancer images in this proposed system has been performed based on the shape and texture characteristics of the images. Thus, we extract two kinds of features: shape and texture. The asymmetry, roundness, intensity levels and more are the real shape and texture features that distinguish the two types of breast tumors. Image processing techniques are used in order to detect tumor and extract the region of interest from the mammogram. The following data processing operations have been done for the extraction of tumors: Thresholding, filtering, adjustments, canny edge detection, and some morphological operations. Texture features are then extracted using GLCM algorithm, while the shape features are extracted directly from the images. The experimental results show a great identification rate of 92%.

***Index Terms*— breast cancer, malignant tumor, benign, texture, canny edge detection, morphological operations, GLCM**

I. Introduction

The breast cancer is about the most common types of cancer among women worldwide and second most common one among women in South Africa, according to the Cancer Association of South Africa according to World Health Organization [1]. Breast cancer is also the top cancer in women in both the developed and the developing world.

Breast cancer is a dangerous medical condition needs to be diagnosed and early detected in order to prevent its growth and reduce the percent-age of deaths caused by it [2].

Breast cancer screening can be achieved using different imaging techniques. The most common screening technique is the mammography. This kind of imaging technique is a specific form of radiography that uses radiation lower than those of conventional radiography such as routine x-ray [3].

specific form of radiography that uses radiation lower than those of conventional radiography such as routine x-ray [3].

In order to come out with a new and unique intelligent breast cancer identification approach, there must be a review of the previous work related to this topic. A proposed method for breast cancer detection was presented in [4] using thresholding and tracking to identify the breast border, but no discussion of the accuracy of the results was presented. The paper described some preliminary works in the analysis of asymmetries in digitized mammograms. They proposed a method for enhancing the asymmetries. The method is to first register, and then bilaterally subtract two mammograms of the left and right breast side in the medio-lateral view. Then, these asymmetries are analyzed in order to provide a tool for computer aided diagnosis (CAD). Another system is proposed in [5] for the identification of the breast edges using areas enclosed by the ISO - intensity contours. The authors used different image processing techniques in order to identify the breast cancer in a mammogram. Such techniques are first thresholding which involves selecting a single gray-level from an analysis of the gray-level histogram, and then segment the mammogram into the background and breast tissue in order to extract the region of interest. Other authors in [6] proposed a methodology utilizing Twin Support Vector Machine (TW-SVM) for the computerized identification of masses in advanced mammograms. The proposed system was assessed by a data set of 100 mammograms obtained from the Digital Database for Screening Mammography (DDSM) database. The outcomes demonstrated that the sensitivity could achieve 89.7% with 0.31 false positive every image. Further examination demonstrated that the proposed CAD framework attained to 94% sensitivity for threatening masses in the test sets, however the detection rate for benign masses was much lower, just 78%.

The aim of this paper is the design of a breast cancer identification system based on the extraction of both texture and shape characteristics of the breast images. It is a part of the ongoing currently conducted researches for detecting and classifying breast tumors, for the purpose of reducing the rate of occurrence of that disease. Moreover, to detect it in its earlier stages, in order to treat it prior to its growth and development. Nevertheless, the proposed work aims to use different and additional methods to reach the desired purpose: detecting breast tumor and classifying it into two main classes: Benign, and Malignant. Hence, the proposed system is based on the combination of image processing techniques and artificial neural networks. Different image processing techniques such as image filtering using median filters, image adjustment, image thresholding, and some morphologi-

Manuscript received April 2, 2016; revised June 23, 2016. This work was supported by the Centre of Excellence in Near East University.

Abdulkader Helwan is a PHD candidate in the Near East University, Biomedical Engineering department. He is also a member of the Applied Artificial Intelligence Research Centre (e-mail: abdulkader.helwan@neu.edu.tr).

Rahib Abiyev is the chairman of the computer engineering department in the Near East University. Prof. Abiyev is the Founder of the Applied Artificial Intelligence Research Centre (e-mail: rahib.abiyev@neu.edu.tr).

cal techniques (erosion, image opening) are used. The shape and texture features (Fig.1) of the breast images are then extracted and fed into a neural network to be classified into benign and malignant tumors.

The structure of the paper is as follows: section one is an introduction about the proposed work, section two is the methodology of the proposed system, in which we discuss the two main phases of it; image processing phase and neural network phase. Finally, section three is a conclusion of the proposed work.

II. IBCIS: METHODOLOGY

In this paper, an intelligent breast cancer identification system (IBCIS) is developed. The system is implemented using Matlab programing language (Matlab 2013 software tools). IBCIS is based on different image processing techniques used in order to segment and extract the breast tumor. The breast images are obtained from DDSM [7]; a public database available on the internet. The images are of size 221*358 pixels. They are processed and rescaled to 256*256 pixels for the purpose of fast and easy computing. Shape and texture features are then extracted from images, in order to be fed into the neural network that has the capability of classifying them into benign and malignant due to its experience gained during the training phase using the backpropagation learning algorithm. Fig. 2 shows the two different phases of the proposed system.

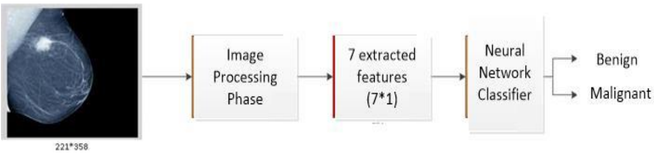

Fig. 1 The different phases of the proposed system

A. Image Processing Phase

The images are first converted to grayscale using the luminosity method. Then, they are filtered and adjusted to increase their pixels intensity, so that the tumor area can be clearer and brighter. The images undergo threshold computing for the purpose of segmenting the region of interest (tumor) located in the breast. We also used some morphological techniques such as dilation and opening in order then to extract the region of interest. The 7 features are then extracted from the region of interest image such as standard deviation, mean, asymmetry, roundness, uniformity etc... Fig. 3 shows the flowchart of the developed identification system. It lists the methods used in the system. Fig. 4 represents an abnormal breast image that undergoes all the processes discussed. Fig. 5 illustrates a normal mammogram that undergoes all discussed processes till the extraction of its region of interest.

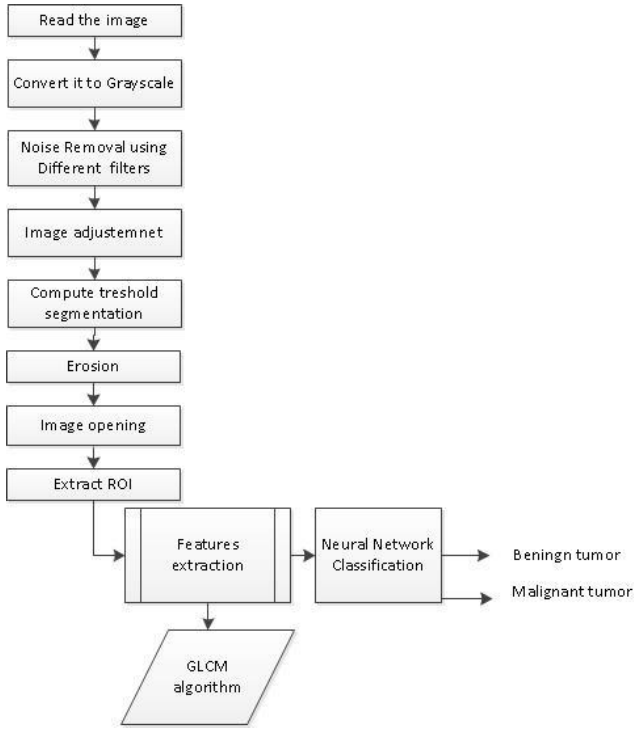

Fig. 2 Flowchart of the proposed identification system

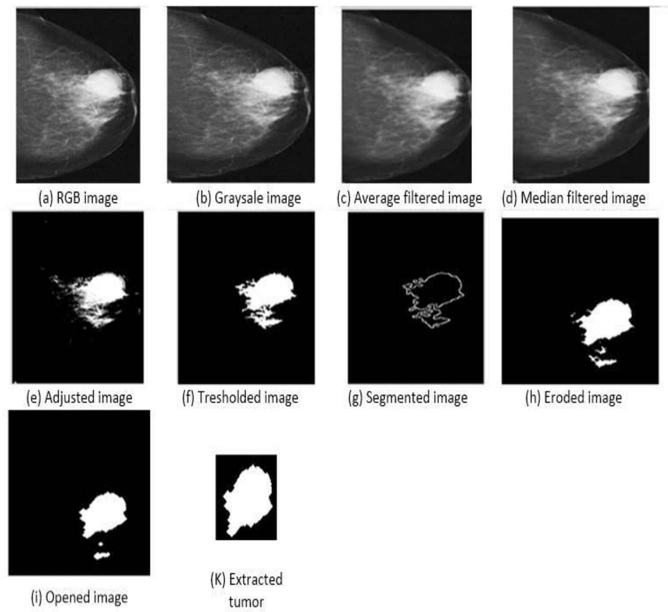

Fig. 3 Malignant breast tumor image undergoes the proposed system algorithm

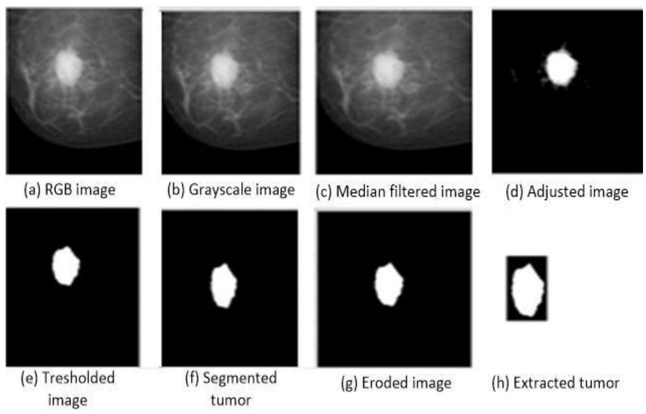

Fig. 4 Benign tumor breast image undergoes the proposed system algorithm

1. Grayscale conversion: The first step is to convert the RGB image to grayscale. This conversion is done using the luminosity method which relies on the contribution of each color of the three RGB colors. Using this method, the grayscale image is brighter since the colors are weighted according to their contribution in the RGB image not ave-ragely [8].
2. Median filtering: The most common types of filters for smoothing purposes are the linear filters such as the me-dian filter which is used in our proposed system. This filter is used to reduce impulsive noise or the salt-and pepper in an image with preserving the useful features and image edges. It is a linear process in which the output of the being processed pixel is found by calculating the median of a window of pix-els that surrounds that studied pixel [9].
3. Image adjusting: The breast images undergo inten-sity adjustment in which the input image's intensities are mapped to a new range of in the output image. This can be done by setting the low and high input intensity values that should be mapped and the scale over which they should be mapped [10].

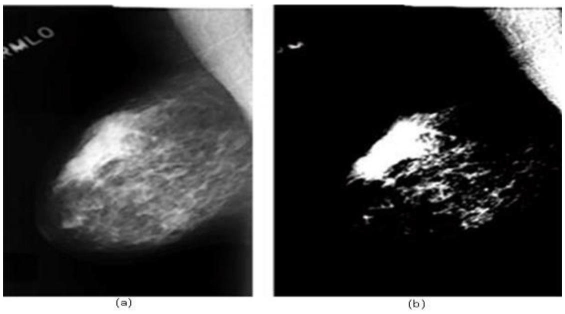

Fig. 5 Image adjusting. (a) original image, (b) adjusted image

4. Thresholding: It is the separation of region of im-ages into two regions. One region corresponds to the fore-ground region, in which it contains the objects that we are interested in. The other region is the background, corres-ponds to the unneeded objects. This provides segmentation of the image based on the image different intensities and intensity discontinuities in the foreground and background regions [10]. The input of this method is usually a grayscale or color image, while the output is a binary image representing the segmentation. The black pixels refer to background and white pixels refer to foreground. The seg-mentation is achieved by a single parameter known as the intensity threshold. This is set by analyzing the histogram of the image which represents the intensity distributions of the image (Fig. 4.f). During thresholding, each pixel is compared to that threshold value of 0.42 in our system. If the pixel value is greater than that threshold, then this pixel is consi-dered as foreground pixel (white). If the pixel value is lower than that threshold value, then the pixel is considered as background pixel (black) [11].
5. Morphological techniques: These can be defined as a set of image processing operations that process images based on shapes. These operations can be done by applying a structuring element in an input image, resulting in an output image of the same size. The structure element is a matrix consists of 0's and 1's, where the 1's are called the neigh-bors. The value of each pixel in the output image is set ac-cording to a comparison of the corresponding pixel in the input image with its neighbors. Structure element has many shapes according to its application. Here, the "disk" structure element with a "radius" of 15 is used. The most common morphological operations are dilation and erosion. The latter is used to shrink the objects in a binary image. After erosion, the main pixels that remain are those that fit altogether with the structuring elements in the foreground view [9]. The thinning of objects is controlled by a little structuring com-ponent of radius. In the accompanying (fig.4.g) you can see the result of erosion an image with a "disk" structure element of "radius" 15 [12].

Image opening was also used in our developed system. This morphological technique is considered as erosion fol-lowed by dilation and it is generally used to smooth the con-tour of an image and break the thin holes (Fig. 4.h) [12].

B. *Feature Extraction*

During this phase, the texture and shape features are ex-tracted in order to discriminate the benign and malignant breast tumors. Since the two different classes (normal, be-nign, and malignant tumor) differ in intensities and shapes; therefore, the two types of features should be extracted from the segmented region of interest (ROI) for the purpose of obtaining accurate classification results. A GLCM is first generated from the segmented ROI; which is a graylevel co-occurrence matrix describes the composition of an image by calculating how frequently a pair of pixels with particular qualities and in a tagged spatial relationship occurs in an image [13]. The texture features are then extracted from the GLCM; however, the shape features are extracted directly from the extracted tumor.

Table 1. Extracted Texture and Shape Features

Features	**Feature number**
Roundness	1
Uniformity	2
Asymmetry	3
Compactness	4
Entropy	5
Standard deviation	6
Mean	7

i. Texture features

Texture features with taking into account the pixels making up the segmented area have been utilized broadly by numerous scientists as a part of the field of advanced mammography [14]. These features can be derived specifically from the pixels estimations of the segmented ROI as indicated by a particular recipe for each feature, or can be computed by implication as far as histogram of the segmented ROI. The following are the formulas of the measures extracted from the segmented ROI of the two classes [14]: benign and malignant breast tumor.

- Mean (average): it is the average intensity of the image. Concerning mammograms, the denser tissue is, the higher the average intensity.

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} p(i,j) \tag{1}$$

Where p(i,j) represents the pixel value at point (i,j), in an ROI of size M×N.

ISBN: 978-988-14048-2-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

- Standard deviation: it can be defined as a measure of the contrast intensity grows, according to the irregularity of the texture.

$$\sigma = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(p(i,j)-\mu)^2} \quad (2)$$

Where p(i,j) is the pixel value at specific point (i,j), and in an ROI of size M×N. μ is the average intensity.

Entropy: it is defined as a disorder, where in the case of texture analysis is a measure of its spatial disorder.

$$h = -\sum_{k=0}^{L-1} Pr_k(\log_2 Pr_k) \quad (3)$$

Where P_{rk} represents the probability of the k-th grey level, and L is the total number of the available grey levels in an ROI of size M×N.

- Uniformity: it is denoted as U and it is a texture measure that is based on the histogram of the segmented ROI.

$$U = \sum_{k=0}^{L-1} pr_k^2 \quad (4)$$

Where P_{rk} is the probability of the k-th grey level. Because the P_{rk} has values in the range of (0 to 1) and their sum equals 1, U is maximum when the numbers of pixels in all grey levels are equal, resulting in all the gray levels to be equal probable and their distribution to be uniform, and decreases otherwise.

ii. Shape features

The shape has a vital role in distinguishing the two different classes of breast tumor. The benign breast tumor is usually a circular and symmetric shape. However, the malignant breast tumor has a random and asymmetric shape [15].

- Roundness: it is the gray level variation in a gray level co-occurrence matrix.

$$R = \frac{4A\pi}{P^2} \quad (5)$$

Where A is the area of the segmented region of interest and P is its perimeter. If the Roundness is greater than 0.90 then, the object is circular in shape.

- Asymmetry: it is to evaluate whether the intensity levels tend to the dark side or light around the mean.

$$A = \sqrt{(x_{ij}-\mu)^2 p(x_{ij})} \quad (6)$$

Where x_{ij} is pixel value at point (i,j), and μ is the mean. The $p(x_{ij})$ is probability of occurrence of that pixel value.

- Shape or Compactness: Since the shape of the segmented ROI is one of the important features that distinguish the benign and malignant tumors, shape features are extracted from each ROI prior to classification.

$$C = \frac{P^2}{4\pi A} \quad (7)$$

Where P is the perimeter, A is the area of the segmented ROI in pixels. The 4π factor is added to the denominator such that the compactness of a complete circle is 1 [15].

C. *Classification phase: neural network*

In this paper, we propose a new approach for the intelligent classification of breast cancer using based on some extracted shape and texture features using image processing techniques and a neural classifier.

The extracted shape and texture features are fed into a neural network that classifies the images into benign or malignant tumor. During this phase, the x-ray images of knee are classified using a supervised backpropagation neural network due to its simplicity and the sufficient number of images. Fig.7 illustrates the topology of the backpropagation neural network used in the proposed system. It consists of three layers: input layer, hidden layer, and output layer. The input layer contains 7 neurons since the extracted features are 7, however; the hidden layer contains 20 neurons, which guarantees significant training while keeping the time expense to a minimum. The output layer consists of two neurons; one represents the benign tumor and one for the malignant cancer.

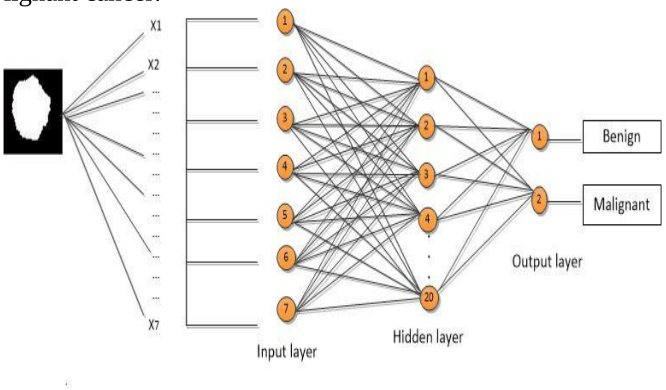

Fig. 6 ANN topology

During this learning stage, initial arbitrary weights of values in the middle of –0.1 and 0.1 were utilized. The learning rate furthermore, the momentum rate; were set through different investigations keeping in mind the end goal to attain to the required minimum error value. A minimum error of 0.001 since it is a medical application. Table 2 demonstrates the final parameters of the trained neural network.

Table 2. ANN input parameters

Parameters	Value
Number of neurons in input layer	7
Number of neurons in output layer	3
Number of neurons in hidden layer	20
Iteration number	5000
Learning rate	0.001
Momentum rate	0.5
Error	0.001
Training time (sec)	300
Activation Function	Sigmoid

ISBN: 978-988-14048-2-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

- System training

The training is done using backpropagation learning algorithm with both adaptive learning rate and momentum, with the function 'traingd'. After making sure that the error is minimized; we started feeding the neural network with the input shape and texture features extracted from the ROI and their targets respectively.

The network was trained on 200 breast images obtained from the DDSM database: 100 images for benign tumor and 100 for malignant tumor (cancer). The table 3 represents the training set of images which consists of two types of images: benign, and malignant. It also shows the total number of database breast images used for training and testing phase. The training is done using backpropagation learning algorithm with both adaptive learning rate and momentum, with the function 'traingd'. After making sure that the error is minimized; we started feeding the neural network with the input shape and texture features extracted from the extracted tumor and their targets respectively.

Table 3. Training and testing number of images

	Benign tumor images	**Malignant tumor images**	**Total number of images**
Training	150	150	300
Testing	150	150	300
Total	300	300	600

Fig.7 shows the error versus iterations graph during system training phase. The network was trained on a large number of images obtained from DDSM. Thus, the result was fine and the network was well trained as the mean square error was diminishing inversely proportional to the increase of the number of epochs.

Table 4. Intervals of ROI extracted feature

Extracted features	**Benign**	**Malignant**
Roundness	0.85-0.99	0.25-0.84
Uniformity	0.98-1	0.81-0.89
Asymmetry	0.90-0.99	0.2-0.89
Compactness	0.2-0.7	0.72-1
Entropy	0.0503-0.304	0.347-0.593
Standard deviation	0.040565-0.238626	0.232185-0.439165
Mean	0.46-0.54	0.556-9

Table 4 shows the intervals of the extracted features of both benign and malignant tumors. These ranges are obtained by finding the minimum and maximum of each feature value. Thus, the range of each feature is between its lowest and highest value.

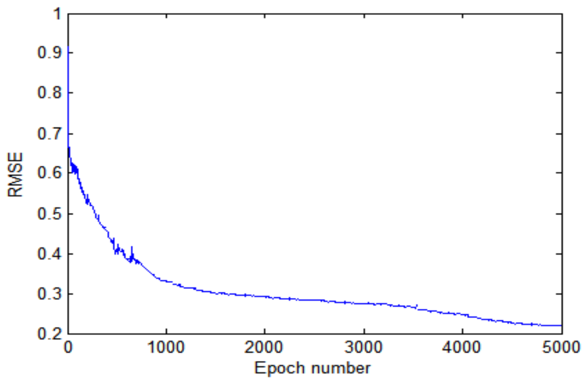

Fig. 7 Learning curve

- System Performance

The proposed breast cancer identification system was implemented using 2.7 GHz PC with 4 GB of RAM, Windows 7 OS and Matlab 2013a software tools. The network was tested on a dataset of 200 images; 100 for a benign tumor and 100 for malignant tumor images. Table 4 shows the results obtained from during training and testing phase of the two different classes (Benign, malignant). It represents the number of images that were accurately recognized by the network in the training and the testing phase. It also shows the percentage of images that were not recognized during the testing phase.

The number of recognized images was divided by the total number of images with respect to each case set (benign and malignant tumor). The result of this fraction is called the classification rate, which is the efficiency of the neural network in classifying the breast cancer.

The experimental results of the intelligent breast cancer identification system were as follows: 100% using the training image set (200 images, 100 for each class). The overall identification rate was eventually calculated and the result is approximately 92% as identification rate. Table 5 shows the intelligent breast cancer identification results in details.

Table 5. Breast cancer identification results

Breast tumors	**Image sets**	**Number of images**	**Identification rate**
Benign tumor	Training set	100	100/100 100%
	Testing set	100	96/100 94%
Malignant tumor	Training set	100	100/100 100%
	Testing set	100	94/100 90%
Total identification rate	All data	400	370/400 92%

III. Conclusion

In this research, an intelligent breast cancer identification system is developed. This novel approach is based on some image processing techniques used to extract the tumor area considering its significant features. Seven features that represent the texture and shape characteristics of the tumor are then extracted and fed into a backpropagation neural classifier. During the network's adaptive learning using backpropagation algorithm, it converges and owns the capability of distinguishing between the benign and malignant breast tumors based on the texture and shape extracted features. In conclusion, it is noticed that the extraction of the texture and shape features has a great effect on the classification phase as it speeds up the network's capability of learning.

References

[1] S. Dehghani, and M. Dezfooli, "Breast Cancer Diagnosis System Based on Contourlet Analysis and Support Vector Machine", World Applied Sciences Journal, vol. 13, No. 5, pp. 1067-1076, 2011.

[2] N. E. B. P. L. F. Lucchini, F. and C. La Vecchia, "Mortality from major cancer sites in the european union, 1955-1998." Annals of Oncology, vol. 14, pp. 490–495, 2003.

[3] T. E. Anders Lernevall, K. and A. Taube, "Benefit of independent double reading in a population-based mammography screening program." Radiology, vol. 191, pp.241–244, 1994.

[4] N. Rederic, F. Faizon, and Y. Sun, "Computerized Analysis of Asymmetry in Digitized Mammograms," in Proceedings of the IEEE 26th Annual Northeast Bioengineering Conference, 2000, pp. 45–50.

[5] J. Padayachee, M. Alport, and W. Rae, “Identification of The Breast Edge Using Areas Enclosed by Iso-intensity Contours,” Computerized Medical Imaging and Graphics, 31, p.p.390–400, 2007.

[6] S. Xiong and L. Jing, “Mass Detection in Digital Mammograms Using Twin Support Vector Machine-based CAD System,” in International Conference on Information Engineering, 2009.

[7] M. Heath, K. Bowyer, D. Kopans, M. Moore, and W. Kegelmeyer, “The Digital Database for Screening Mammography,” In Proceedings of the Fifth International Workshop on Digital Mammography, pp. 212–218, 2001.

[8] Z. Wang and D.Zhang, "Progressive switching median filter for the removal of impulse noise from highly corrupted images", IEEE Trans. on circuits and Systems II: Analog and Digital signal processing , Vol. 46, no. I , pp 78-80, 1999..

[9] H. Abdulkader, “ITDS: Iris Tumor Detection System Using Image Processing Techniques,” International Journal of Science and Engineering Research, p.p.45-80, 2014.

[10] R. Gonzalez, and E. Woods, Digital Image Processing, 2cd ed. Prentice-Hall, 2002.

[11] A. Helwan and R. H. Abiyev, "ISIBC: An intelligent system for identification of breast cancer," 2015 International Conference on Advances in Biomedical Engineering (ICABME), Beirut, 2015, pp. 17-20. doi: 10.1109/ICABME.2015.7323240

[12] S.S. Mokri, M.I. Saripan, M.H. Marhaban, A.J. Nordin, “Lung segmentation in CT for thoracic PET-CT registration through visual study,” in Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on, Langkawi, p.p. 550 - 554, 17-19, Dec. 2012.

[13] L. ZHU and Z. ZHANG, “Auto-classification of Insect Images Based on Color Histogram and GLCM,” in the Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Yantai, Shandong, p.p. 2589 – 2593, 2010.

[14] R.M. Haralick, K.S., I. Dinstein, "Textural features for image classification". IEEE Trans. Syst. Man Cybern. 3(6): p. 610–621, 1973.

[15] D. Clausi, and Y. Zhao, “Rapid Co-occurrence Texture Feature Extraction Using a Hybrid Data Structure,” Computers and Geosciences, *28*, p.p.763–774, 2002.