COPMUTER-AIDED DIAGNOSTIC SYSTEM FOR MASS DETECTION IN DIGITIZED MAMMOGRAMS

I. M. Ibrahim, A. A. Yassen, A. F. Qurany, G. E. Essam, M. A. Hefnawy, M. A. Yacoub, Y. M. Kadah
Systems and Biomedical Engineering, Cairo University, Giza, Egypt
e-mail: i.m.ibrahim@k-space.org

Abstract-**Physician experience of detecting breast cancer can be assisted by using some computerized feature extraction algorithms. In this study, we propose a system that extracts some features from the breast tissue digital mammogram image. Then, the discrimination power of these features is tested to avoid using non-classifying features in order to minimize the classification error. The feature extraction step was applied over 102 images coming from 20 cases. These images are divided into two independent sets; the learning set and the testing set. Features from the first set are further used to learn the system how to differentiate between normal and cancerous breast tissues. The testing set is used to test the power and the accuracy of the system. Two statistical classifiers were used and compared through the system to reach a better classification decision. Changing the window size and the overlapping volume through extracting the features is studies also. The best results gave a sensitivity of 75 % and a specificity of 71.4 %.**

***Keywords*- Computer-aided diagnosis, mammography, feature extraction, statistical classifiers**

I. INTRODUCTION

Brest cancer is one of the most important causes that contribute to mortality in women. The earlier the cancer is detected, the higher the chance of survival for patients. Mammography is the most effective method that is used in the early detection of breast cancer [1], [2]. Masses are one of the signs that have to be detected in mammograms. Retrospective studies showed that radiologists can not detect all the masses in the mammograms. Some reasons of this misdetection refer to human factors such as decision criteria, simple oversight, and distraction by other image features. These errors may occur with experienced radiologists [2]. Ciatto et al. [3] showed that a computer-aided detection system (CAD) can help radiologists in taking their decision about detecting tumors in the mammograms.

Many techniques have been used to detect masses in the mammograms. Youssry et al. [4] used a technique that depends mainly on the difference between normal and cancerous histograms and used four features for the classification process through a neural network classifier. The four features are statistical ones which are the mean and the first three moments. Preprocessing techniques were used such as histogram equalization and segmentation. Yu *et al.* [1] presented a CAD system for the automatic detection of clustered microcalcifications through two steps. The first one is to segment potential microcalcification pixels by using wavelet and gray level statistical features and to connect them into potential individual microcalcification objects. The second step is to check these potential objects by using 31 statistical features. Neural network classifiers were used. Results are satisfactory but not highly guaranteed because the learning set was used in the testing set.

Fogela *et al.* [5] used the patient age as a feature besides radiographic features to train artificial neural networks to detect breast cancer. Verma *et al.* [6] presented a system based on fuzzy-neural and feature extraction techniques. A fuzzy technique in conjunction with three features was used to detect a microcalcification pattern and a neural network to classify it into benign or malignant. Brake et al. [7] studied the scale effect on the detection process by using single scale and multi-scale detection algorithms of masses in digital mammograms.

In our study, we propose a CAD system for detecting masses in the digitized mammograms. This study is done through two main phases; the learning phase and the testing phase. This is shown in fig. 1. Through the learning phase, we learn the system how to differentiate between normal and cancerous cases by using normal and cancerous images. In the testing phase, we test the performance of the system by entering a test image to compute the correctness degree of the system decision.

This paper is arranged as follows. Section II covers the used methods. Results and discussion are found in Section II while Section IV contains the conclusion of this study.

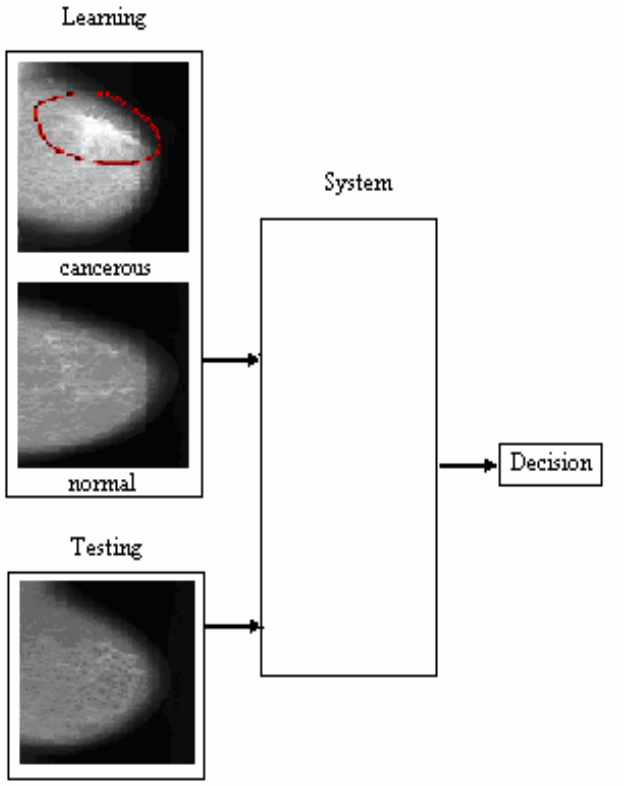

Fig. 1. Block diagram of the proposed system.

II. METHODS

Fig. 2 illustrates the two main phases of the system; the learning phase and the testing phase. Each phase is composed of two major steps. They are the feature extraction step and the classification step. Features resulting from the first phase are followed by the step of feature selection through the t-test. The number of used features before the t-test is 25.

Two statistical classifiers are used; the minimum distance classifier and the voting k-Nearest Neighbor (k-NN) classifier. We compared the results obtained from the two classifiers. Also, we studied the effect of changing the window size and the overlapping area on the results.

In this study, we did not use preprocessing techniques such as smoothing, edge sharpening, or wavelet decomposition. We just dealt with the mammograms as raw data without any alteration in it. This is because we do not know exactly what the underlying data is. So, we could not choose an enhancement technique for not being biased to a wrong one. Also, we wanted to test the performance of our system on the data as it is.

A. Feature Extraction

We used 25 features. 22 of them are conventional features and 3 are unconventional features that were used in other studies and they showed good results. The 22 conventional features are mean, standard deviation, variance, skewness, kurtosis [8], and entropy [9]. Nine percentile features were used ranging from the first percentile up to the ninth percentile [8]. Also we used the seven invariant moments that are invariant to scale, translation, and rotation change [10].

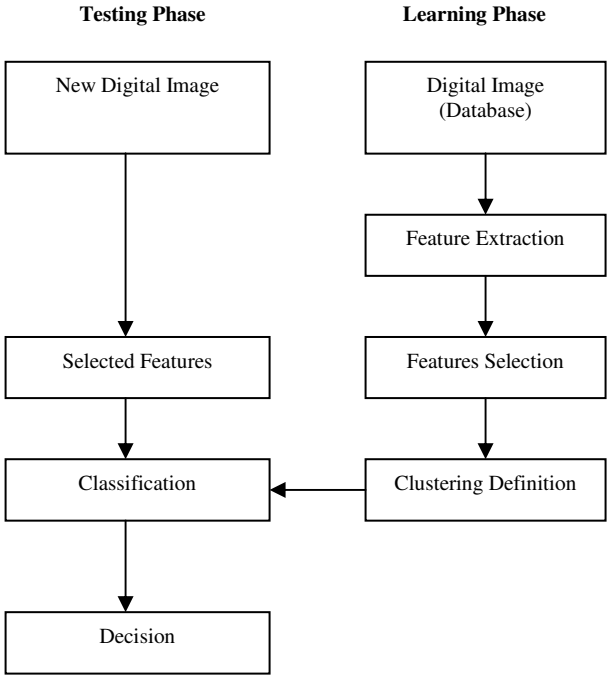

Fig. 2. Our system for mass detection in digitized mammograms.

The three unconventional features are the median contrast, the normalized gray level value [1], and the spreadness [11]. They are described as follows:

$$c(i,j) = p(i,j) - \text{median}\left(y(l,m), l,m \in \text{Window}\right) \quad (1)$$

$$s(i,j) = \frac{p(i,j) - \text{mean}\left(y(l,m) : l,m \in \text{Window}\right)}{\text{std}\left(y(l,m) : l,m \in \text{Window}\right)} \quad (2)$$

$$f = \frac{\sum_i \sum_j p(i,j)(i-i_0)^2 + \sum_i \sum_j p(i,j)(j-j_0)^2}{\sum_i \sum_j p(i,j)} \quad (3)$$

where $p(i,j)$ is the pixel value at position (i,j), Window is an $n \times n$ square area centred at position (i,j), std is the standard deviation of the pixel values in the Window, (i_0, j_0) are the coordinates values of the centred pixel, $c(i,j)$ is the median contrast at position (i,j), $s(i,j)$ is the normalized gray level value at position (i,j), and f is the spreadness.

We applied the previous features on our images. The normal images were of size 520 x 500 while the cancerous images were of variable size due to the size of the cancer in each case. We moved over these normal and cancerous regions of interest (ROI) with a window size of 64 x 64 pixels and an overlapping shift of 32 x 32 pixels. We chose these sizes as moderate size in computations and we studied changing them also as will come next. The output of this step is matrix for each image. Each element in this matrix represents the feature value at a certain position of the window through the ROI. These matrices are used in the t-test as follows.

B. t-test

The purpose of this step is to get the features that have the ability of differentiation between normality and cancer to be used in the classification process. In other words, we test the discrimination power of the features. The input to this test is two sets of values for each feature. One set represents the normal case and the other set represents the cancerous case. We assume that each set follows a t distribution. The t-test checks the amount of overlapping between the two distributions. If there is no overlapping, then this feature has the ability of differentiation. But in nature, it is not easy to find complete independent distributions without overlapping. So, we determine a significance level to consider the two sets come from two different distributions. We chose this significance level to be 5 %. It means that the probability of incorrectly considering two independent distributions is 0.05 while the truth is that the two sets come from the same distributions. The test computes a value called the p-value which is the probability of observing one sample from the first set in the second distribution. If the p-value is less than the significance level, then these two sets come from two different distributions and this feature can differentiate [8].

To prepare the two sets of each feature, we used the feature matrix resulted from the step of features selection. For each feature, we transfer the matrix of each image to a vector. Thus, we have for each feature a number of vectors equal to the numbers of the sample normal and cancerous images. These vectors are concatenated under each other to form the normal cluster and the cancerous cluster as show in fig. 3. These two sets are the input to the t-test step.

The previous process was done for the 25 features to test their discrimination power to avoid using non-classifying features to reduce the classification error.

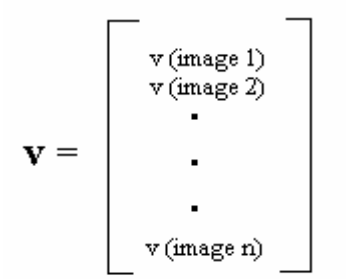

Fig. 3. Forming the feature cluster vector.

C. Classification

Here, we used two statistical classifiers; the minimum distance classifier and the voting k-Nearest Neighbor (k-NN) classifier [9]. We classified the images by the features resulting from the t-test that have the discrimination power.

1) Minimum distance classifier: The distance is the norm of a vector of size M x 1; where M is the number of classifying features resulting from the t-test. Here, we get the mean value of each cluster by getting the average value of the vector representing the whole images. This vector (V) is the one described in fig. 2. For the test sample, we compute the M features and put them in a vector. Then, we compute the distance between this last vector and the two vectors representing the two clusters; normal and cancerous. We assign the test sample to the nearest cluster.

2) Voting k-Nearest Neighbor (k-NN) classifier: The features of the sample images forming each cluster are not concatenated under each other. Instead, they are left separately through the cluster. For the test image, we calculate the features vector of size M x 1. Then, we get the distance between this vector and every sample image in the two clusters. After that, we sort these distances in ascending order. With the choice of k, we assign the test sample to its class. The value of k must be odd. If $k = 1$, the first distance is the smallest one and we classify the test sample to be from the cluster having the learning sample of the minimum distance. With $k = 3$, the test sample is classified to be belonging to the cluster that has 2 or 3 distances from the minimum 3 distances in the ascending vector. Through this study, we compared the results of using $k = 1$ and $k = 3$.

D. Changing the Window Size and the Overlapping Amount

Through the previous work, we were traversing the ROI with a window size of 64 x 64 pixels and an overlapping amount of 32 x 32 pixels. We wanted to study the effect of changing these two parameters. So, we fixed the window size and changed the overlapping amount from 48 x 48 pixels to no overlapping. Also, we fixed the overlapping parameter and changed the window size to 48 x 48 pixels and 80 x 80 pixels. This part of the study was applied only on the highest 3 discriminating features due to problems in time. The whole previous work was repeated using these 3 features only. These 3 features are the mean, standard deviation, and the entropy.

III. Results and Discussion

A. Database

We used digital mammograms from a database called Digital Database for Screening Mammography (DDSM). This is found on the University of South Florida Digital Mammography home page. The DDSM is a resource for use by the mammographic image analysis research community. Primary support for this project was a grant from the Breast Cancer Research Program of the U.S. Army Medical Research and Materiel Command [12]. We used 20 cases divided into two sets; the learning set and the testing set. The learning set is composed of 30 cancerous images and 52 normal images while the testing set contained 8 cancerous images and 14 normal ones. The normal images are taken from the same image that has cancerous regions. The all cases come from the same digitizer which is lumisys. We chose the size of the normal images to be 520 x 500 pixels while the size of the cancerous images was determined according to a file informing the cancer position in each case.

B. Feature Selection

The t-test resulted in 20 features that can differentiate between cancer and normality. The excluded features are shown in Table I. They are excluded because their p-value is larger than the significance level which was set to 5 %.

C. Classifiers Results

For each classifier, we calculated the sensitivity and the specificity. Sensitivity is the conditional probability of detecting cancer while there is really cancer in the image. Specificity is the conditional probability of detecting normal breast while the true state of the breast is normal. Results of Table II are those of the minimum distance classifier. Table III shows the results of the voting k-NN classifier with varying the value of k to take 1 and 3. The minimum distance classifier gives the best results. The voting k-NN classifier with $k = 1$ is better than that of $k = 3$ but both are worse than the minimum distance classifier.

The most important factor in judging the performance of any classifier is the sensitivity parameter. This parameter should be high as possible as we can. This parameter means the ability of detecting cancerous cases.

TABLE I
EXCLUDED FEAURES AND THEIR P-VALUES

Feature	p-value
Skewness	0.1
Kurtosis	0.94
Normalized gray level value	1
5th Invariant moment	0.97
6th Invariant moment	0.9

TABLE II
RESULTS OF THE MINIMUM DISTANCE CLACIFIER

Parameter	Learning set	Testing set
Sensitivity	76.67%	75%
Specificity	54.71%	71.43%

TABLE III
RESULTS OF THE VOTING k-NN CLACIFIER

Parameter	k = 1		k = 3	
	Learning set	Testing set	Learning set	Testing set
Sensitivity	100%	50%	90%	37.5%
Specificity	100%	71.43%	88.68%	78.57%

If the case is cancerous and the system failed in detecting it, this will be a life threatening matter. But if the case is normal and the system classified it as cancerous, this error will be fixed by any further investigation like biopsy sample. So, the results of the minimum distance are the best.

These results are not so much satisfactory. This returns to many reasons. The first reason comes from the great variability in the database mammograms. The cancer values and the normality values change extensively which leads to more overlapping between the normal cluster space and the cancerous cluster space. The second reason is the small number of used cases in learning the system which does not cover the entire space of each cluster. The used testing set forms the third reason. Some of these samples are not used in the learning phase. So the system faced difficulty in recognizing something that it does not know as there is no similar case in the learning phase.

Also, these results are not accurate to a great extent due to not fixing one parameter of the study parameters. It is the size of the selected ROIs. The normal images size was fixed to 520 x 500 pixels but for the cancerous images it differed according to the size of the cancer that is determined by the associated file with the case. It was necessary to fix this parameter by taking fixed cancerous ROIs. And in this case we were going to take ROIs of cancer only and other of cancer and normality which was going to be a healthy matter as we do not know the position of the cancer in the new case and the process of taking any region of it for investigation can be of cancer only and can be of cancer and normality.

D. Results of Changing the Window Size and the Shift

Changing the window size or the shift amount did not lead to any change in the results. Results remained as it was. So, the usage of moderate window size and no overlapping can lead to better computation time.

IV. CONCLUSION

In this study, we proposed a system for mass detection in the digitized mammograms of the breast. This system depends on selecting some features and using them in the classification process. We proved that the features of the skewness, kurtosis, normalized gray level value, 5th invariant moment, and the 6th invariant moment can not differentiate between normality and cancer after testing their discrimination power. Also, the minimum distance classifier is better than the k-Nearest Neighbor (k-NN) classifier as the first one gave better results for the sensitivity and also gave close results of the specificity with respect to the (k-NN) classifier. However, caution must be considered while dealing with these results as we used variable cancerous images in the learning phase while the normal images were of fixed size. More cases must be added to the learning set and to the testing set to cover the whole cluster space to obtain better results. The choice of moderate window size is preferable for providing less computation time as this parameter resulted in no change in the results. Also, there is no need for traversing the images with overlapping windowing.

REFRENCES

[1] Songyang Yu and Ling Guan, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Trans. Med. Imag.*, vol. 19, pp. 115-126, February 2000.

[2] Huai Li, K. J. Ray Liu, and Shih-Chung B. Lo, "Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms," *IEEE Trans. Med. Imag.*, vol. 16, pp. 785-798, December 1999.

[3] Stefano Ciatto et al., "Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography," *European Journal of Radiology* 45 (2003) 135–138.

[4] Noha Youssry, Fatma E.Z. Abou-Chadi, Alaa M. El-Sayad, "A neural network approach for mass detection in digitized mammograms," ACBME, 2002.

[5] David B. Fogela, Eugene C. Wasson b, Edward M. Boughtonc, Vincent W. Pm-to, " A step toward computer-assisted mammography using evolutionary programming and neural networks," Cancer Letters 119 (1997) 93-07.

[6] Brijesh Verma and John Zakos, "A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques," *IEEE Trans INFORMATION TECHNOLOGY IN BIOMEDICINE,*, vol. 5, pp. 46-54, March 2001.

[7] Guido M. te Brake and Nico Karssemeijer, "Single and multiscale detection of masses in digital mammograms," *IEEE Trans. Med. Imag.*, vol. 18, pp. 628-639, July 1999.

[8] CHAP T. LE, Introductory Biostatistics. A John Wiley & Sons Publication, April 2003.

[9] Yasser M. Kadah, Aly A farag, Ahmed M. badawy and Abou-Baker M. Youssef, "Classification algorithm for quantitative tissue characterization of diffuse liver disease from ultrasound," *IEEE Trans. Med. Imag.*, vol. 15, pp. 466-478, August 1996.

[10] Rafael Gonzalez and Richard woods, Digital images processing. Prentice Hall, 2002.

[11] Hidefumi Kobatake, Masayuki Murakami, Hideya Takeo, and Sigeru Nawano, "Computerized detection of malignant tumors on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 18, pp. 369-378 May 1999.

[12] http://marathon.csee.usf.edu/Mammography/Database.html