Development of computer aided diagnosis system for mammograms

Gaafar al-aidaros
Student
King abdulaziz university

***Abstract* - to detects and diagnoses breast cancer, mammography is a noninvasive technique has been successful in improving detection of cancer particularly non-palpable breast masses and calcifications that may be malignant. occasionally, radiologists fail to detect suspicious abnormalities that is repetitive and fatiguing task. thus, there is a necessity for developing methods for automatic with more accuracy detection and classification of suspicious areas in mammograms, as a means of helping radiologists to improve the efficacy of screening programs and avert unnecessary biopsies. By incorporating the expert knowledge of radiologists, the computer-based systems provide a second opinion in detecting abnormalities and making diagnostic decisions. such a diagnostic computerized system is called computer aided diagnosis (CAD) system. there are strong motivations to develop a CAD system to aid radiologists in reading mammograms. this paper aimed to develop a computer-aided diagnosis (cad) system by applying new feature extraction and classifiers that affect directly on the CAD system attitude. the experimental results showed better system performance.**

***keywords* - breast cancer, mammography, computer aided diagnosis, feature extraction**

I. INTRODUCTION

Breast cancer is a heterogeneous group of diseases, it is known to start as a local lesion in the breast. It gradually spreads to the lymph nodes in armpits and then distal spread to other organs. In Saudi Arabia, an estimated 2700 new cases of breast cancer annually which it represents 19.9% of cancer cases overall. Therefore, breast cancer ranked first among the most common types of cancer in Saudi Arabia where 43.5% occur in women between 35-54 years old [1]. Throughout the world, there was 571,000 breast cancer deaths in 2015.

Instead of using biopsy test directly that require surgical intervention for examination to determine the presence of a disease, Mammography is the most effective technique that uses low-energy x-rays to examine the human breast and is used as a diagnostic and a screening tool. The goal of mammography is the early detection of breast cancer, typically through detection of characteristic masses and/or microcalcifications [2].

Computer aided tools have been shown to be powerful systems to overcome missed breast lesion problem by physicians when examine a huge number of mammographic images [3]. The sensitivity of readers can be increased by an average of 10% with the assistance of CAD systems. This paper focused on feature extraction that characterize the lesions to differentiate actual lesions from falsely detected candidates, where a set of features is calculated on the selected region of interest (ROI).

II. METHODOLOGY

We used MIAS (Mammographic Image Analysis Society) database that contains left and right breast images for 161 patients. MIAS database consists of 322 images, which belong to three classes such as normal, benign and malignant [4]. MATLAB toolbox (Version of R2020b) was used as a programming environment. We developed the CAD system by following the main next stages that are shown in Fig. (1).

2.1 Preprocessing stage

Starting with preprocessing stage to enhance the peripheral area visibility of the breast that is uncompressed region of the projected breast. In this region, the tissue thickness is smaller than in the interior part of the mammogram [5].

2.2 ROI Extraction stage

We used the information of digital mammographic images from the Mammographic Image Analysis Society (MIAS) database, where there are normal and abnormal mammograms (144 mammographic images). Extraction of 144 centered ROI that was done by using window of size 32x32.

2.3 Feature Extraction stage

In this study, we focused on this stage where it affects directly on the CAD system performance where the extraction of statistical features was from each selected ROI. We used more effective features to get good results [6,7].

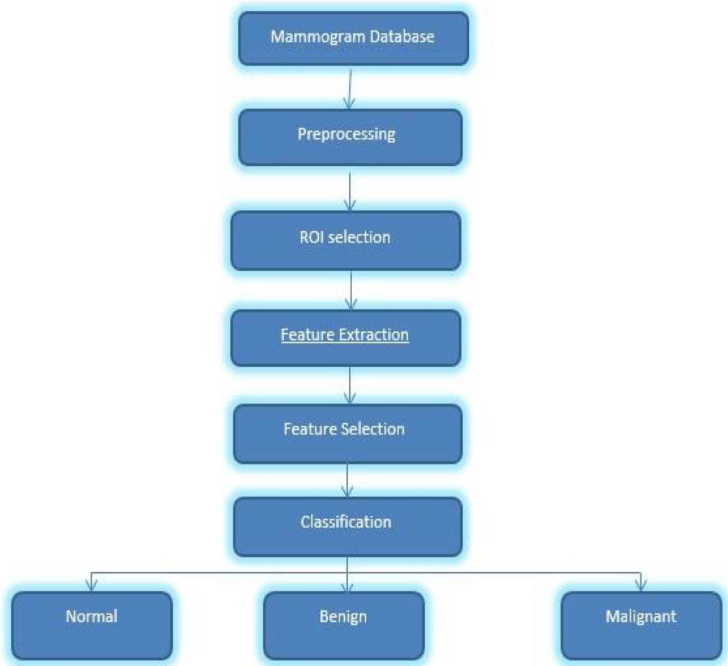

Fig. (1) The schematic diagram for CAD system.

2.4 Feature selection stage

Feature selection is important stage where we chose the most powerful features from the extracted features before. The good choices for features will reduce classification error, then we can obtain robust CAD system. By use MATLAB Statistics Toolbox, we applied some selection methods such as T-test and Sequential Forward Selection (SFS) [8-10].

2.5 Classification stage

This is final stage in the system where the features that are selected from the last stage will pass to the classification phase. the classification is performed using the classifier that clarify the nature of the lesion whether it is a normal or abnormal. In this study, we used K-voting Nearest Neighbor Classifier and Support Vector Machine Classifier [11].

2.6 Evaluation of the system

After classification stage, we evaluated the performance of CAD system by five indices most commonly used in radiologic researches. These indices are sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy.

III. RESULTS

We implemented new features in the first order features that computed from the ROI by changing the histogram levels from 32 to 16. The accuracy increased to 0.9722 as shown below in Table (1).

Table (1) Comparison of performance evaluation with different histogram levels

Performance measures	**The histogram levels=32**	**The histogram levels=16**
Sensitivity (Sens)	0.9444	0.9722
Specificity (Spec)	0.9444	0.9722
Accuracy (Acc)	0.9444	0.9722
Positive Predictive Value (PPV)	0.9444	0.9722
Negative Predictive Value (NPV)	0.9444	0.9722
Area Under the Curve (AUC)	0.9769	0.9807

.

By using the T-Test after changing the histogram levels to 16, the number of useful features (P-Value < 0.05) = 8 out of 11 (Fig. 2). When using Correlation coefficient method, the number of useful features (Correlation Coef.> 0.5) = 6 out of 11 (Fig. 3). The accuracy, Sensitivity, Specify, PPV, NPV, AUC are increasing at levels 16 (Fig. 4).

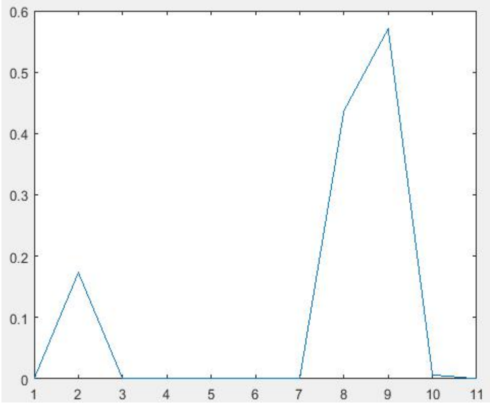

Fig. (2) T-test when histogram levels=16.

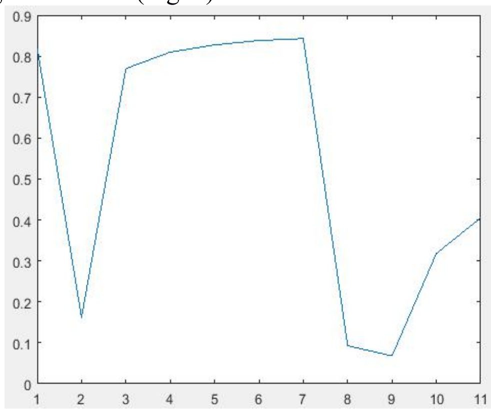

Fig. (3) Correlation Coefficient when histogram levels=16.

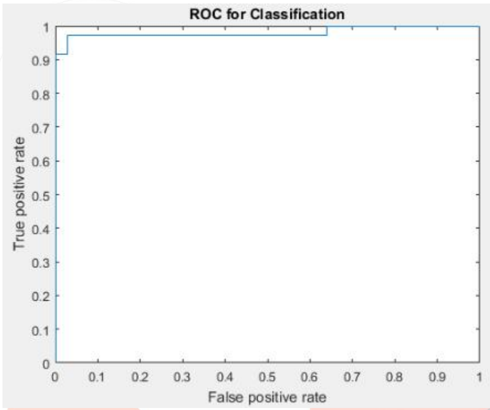

Fig. (4) Receiver operator characteristic when histogram levels=16.

In the classification stage, we used three classifiers that were First Number neighbors classifier (C3), Three Number neighbors classifier(C4) and Gaussian or Radial Basis Function Kernel classifier (C8). From the next Table 4.8, we observed that the best classifier is C8 (Gaussian or Radial Basis Function Kernel classifier) because it has the highest Sensitivity, specificity, Negative Predictive Value (NPV) and accuracy also it has the same Positive Predictive Value (PPV) and Area Under the Curve (AUC). The disease prevalence for this classifier was the lowest value. T-test, correlation coefficient and Receiver Operator Characteristic (ROC) is shown in Table 2 for the three classifiers selected.

Table (2) Comparison of performance evaluation for three classifiers selected

Performance measures	**C3**	**C4**	**C8**
Sensitivity (Sens)	0.8537	0.9459	1
Specificity (Spec)	0.9677	0.9714	0.973
Positive Predictive Value (PPV)	0.9722	0.9722	0.9722
Negative Predictive Value (NPV)	0.8333	0.9444	1
Accuracy (Acc)	0.9028	0.9583	0.9861

IV. CONCLUSION

The selected ROIs from MAIS database divided into two classes (36 images for learning and 36 images for testing). The proposed histogram level was 16 for first order features which made the accuracy increased to 97%. The most effective classifier was Radial Basis Function Kernel classifier when we compared it with First Number neighbors classifier and Three Number neighbors classifier. The accuracy was 98%, the sensitivity was 100% and negative predictive value was 100%.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. So 1. Al.-Madouj A, Alshahrani Z, Alrawaji A, Hayder M, Al-Shridah M, Al-

Shamrani T, Cancer Incidence Report Saudi Arabia 2013.http://www.chs.gov.sa/Ar/HealthCenters/NCC/CancerRegistry/CancerRegistryReports/2013.pdf.

[2] Yu, Guan, "A Cad System For The Automatic Detection Of Clustered Microcalcifications In Digitized Mammogram Films", IEEE Transactions On Medical Imaging, Vol. 19, No. 2, February 2000.

[3] M. P. Sampat, M. K. Markey, A. C. Bovik, "Computer-aided detection and diagnosis in mammography", in Handbook of Image and Video Processing(ed. Bovik), 2nd edition 2005, pgs. 1195-1217.

[4] The mini-MIAS database of mammograms ‖ , Available: http://peipa.essex.ac.uk/info/mias.html.

[5] T Wu, RH Moore, DB Kopans, "Multi-threshold peripheral equalization method and apparatus for digital mammography and breast tomosynthesis ", US Patent 7,764,820, Google Patents, 2010.

[6] K. Bovis, S.Singh, " Classification of mammographic breast density using a combined classifier paradigm" .4th international workshop on digital mammography, pp 177-180, 2002.

[7] A.Bosch, X. Munoz, A. Oliver, and J. Marti, " Modeling and classifying breast tissue density in mammograms" ,Proceedings IEEE computer societyconference on computer vision and pattern recognition, vol 21, pp 1552-1558,2006.

[8] F.J. Massey, " The Kolmogorov- smirnov test for goodnessof fit ", Journal of the American Statistical Association., Vol. 46, No. 253, pp. 6878, 1951.

[9] R.P.W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, and D.M.J. Tax, "A matlab toolbox for pattern recognition", Delft University of Technology, 2004.

[10] Whitney, A.W., "A Direct Method of Nonparametric Measurement Selection", IEEE Transactions in Computers, 1100—1103, 1971.

[11] Y. M. Kadah, A. A. farag, A. M. badawy, and A. M. Youssef, "Classification algorithm for quantitative tissue characterization of diffuse liver disease from ultrasound," IEEE Transactions on Medical Imaging, vol. 15, no. 4, August 1996.