Classification of Breast Mammograms into Benign and Malignant

Muhammad Talha [1, 2,] Ghazali Bin Sulong[2] and Abdulrahman Alarifi[3]

[1]*College of Science Research Centre, King Saud University, Riyadh, Saudi Arabia*
[2]*Department of Computer Graphics and Multimedia, Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, Malaysia*
[3]*Computer Research Institute, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia*
mnaseem@ksu.edu.sa, ghazalis@utm.my, aarifi@kacst.edu.sa

Abstract

In this paper, we have proposed a method that consists of combination of different methods. First we have performed enhancement on breast mammogram to enhance the image quality. After that discrete cosine transform has been applied for features extraction. Bayesian Classifier has been used for classification into benign and malignant. It has been noted that results are very much satisfactory. We have used MIAS data set for experimentation purpose. Proposed method performs good when we have tested on different images.

Keywords: *breast cancer, mammogram, enhancement, bayesian*

1. Introduction

There are a number of well-known and potential risk factors for breast cancer. These can be divided into seven broad categories: age, hormonal factors, family history of breast cancer, proliferate breast disease, irradiation of the breast region at an early age, lifestyle factors and personal history of malignancy [1, 2, 3, 4, 5, 6, 7]. In reality, estimates indicate that between 10 to 30% of breast cancers are missed by radiologists during routine screening. The penalty of errors in detection or classification is very high. Mammography itself cannot prove that a suspicious area is malignant or benign. To decide that, the tissue has to be removed for examination using breast biopsy techniques. A false positive detection may cause an unnecessary biopsy. Statistics show that only 20% to 30% of breast biopsy cases are proved cancerous. In a false negative detection, an actual tumor remains undetected that could lead to higher costs or even to the cost of a human life. With the growth of computer technology, radiologists have a chance to improve their image interpretation using computer capabilities that can improve the image quality of mammograms [8, 9, 10, 11, 12]. In order to develop the accuracy of interpretation, a variety of computer-aided diagnosis (CAD) systems like [13, 14, 15] have been proposed. CAD plays an important role in diagnosis of breast cancer and defining the extent of breast tumors. In previous twenty years, much effort has been made by computer scientists to support the radiologists in detection and diagnosis of cancerous masses by developing computer-aided tools for mammography interpretation. Image processing and intelligent systems are two important mainstreams of computer technologies that have been continuously explored in the development of computer-aided mammography systems.

From computer vision point of view, automatic segmentation and classification of mammogram images is not easy to address. Three types of anatomical variations in the tissues of mammograms image are present from person to person; i.e. fatty, fatty-glandular and dense-glandular. Inherent limitations in the imaging process due to low dose X-Rays which often result in noisy images. Radiologists are trained to differentiate between benign and malignant abnormalities but for computer it needs proper training to adopt accordingly. We have proposed fully automatic and robust technique. Strong preprocessing technique and automatic abnormality type detection method is used. No prior knowledge of the mammogram is required about its feature, type, and contents. This is a supervised method for diagnosing breast cancer. Proposed system achieved quite good accuracy for the classification of mammograms as malignant and benign. Paper is organized as follows: Introduction is given in Section 1. Section 2 discusses related work. Section 3 describes the proposed method. Section 4 includes experimental results and conclusion is presented in section 5.

2. Proposed System

The proposed system is divided into four major parts as shown in Figure 1:

- Enhancement by using Histogram Equalization
- Features Extraction
- Classification

The detail of these four steps is described below one by one.

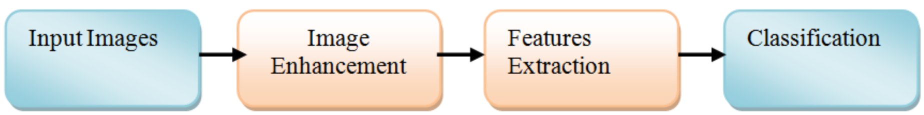


Figure 1: Proposed Method

3.1 Preprocessing for Enhancement

In this step, Histogram Equalization technique has been applied. The enhancement is condensed in flat areas of the image, which prevent over enhancement of noise. It also reduces the edge shadowing effect.

3.2 Features Extraction

Features play a significant role in CAD (Computer Aided Diagnostic) environment. We have used DCT feature for our proposed system.

3.2.1 Discrete Cosine Transform (DCT) Features

Discrete cosine transform (DCT) is used for converting the signal into its frequency components. In image processing DCT attempts to de-correlate the image data [12]. DCT has the ability to pack the image data into as few DCT coefficients as possible without any distortion. DCT has the property of separability and symmetry. 2-Dimensional DCT of the input is defined by the following equation:

$$C(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)cos\left[\frac{\pi(2x+1)u}{2N}\right]cos\left[\frac{\pi(2y+1)v}{2N}\right] \quad (1)$$

Where $0\leq u \leq N$, & $0\leq v \leq N$, and

$$a(u)=\begin{cases}\sqrt{\frac{1}{N}} & for\ u=0\\ \sqrt{\frac{2}{N}} & for\ u\neq 0\end{cases} \tag{2}$$

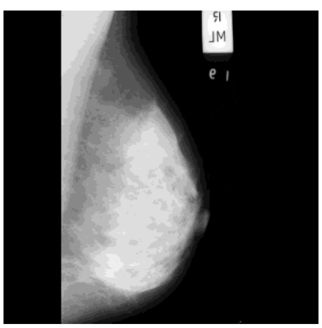
3.3 Classification

For estimating the class of the new data item a probabilistic model is defined which is known as Bayesian classification. Bayesian classifier is a statistical classifier. Bayesian classification is based on the bayes theorem [14]. Bayesian classification is used for classifying objects into associated classes based on the attributes of these objects. Attributes of the data/object are considered as independent of each other in Naive Bayes classification [15, 16].

In Naïve bayes classification dataset is divided into two sets, training and testing respectively. Training dataset is considered as prior information and model is constructed on the basis of this training dataset. Class of the unknown data is determined using this model [16].

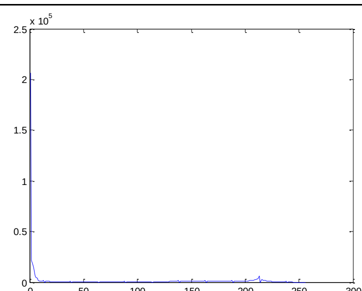
4. Results and Discussion

We have used publically available databases MIAS [17]. The dataset is taken from the Mammographic Institute Society Analysis (MIAS). Each mammogram is of size 1024 x 1024 pixels, and resolution of 200 micron. There are 322 mammograms of right and left breast of 161 patients in this dataset. 69 mammograms were diagnosed as being benign, 54 malignant and 207 normal. Enhancement has been done by histogram equalization. Results have been show in Figure 1.

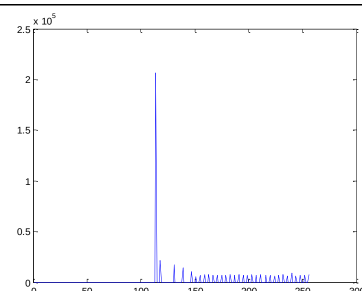


Original Image

Histogram Equalization



Original Image Histogram



Histogram Equalization Histogram

We have tested the performance of the Bayesian classifier by calculating and analysis of accuracy, sensitivity and specificity for malignancy detection. Performance of classifiers is calculated and analyzed by the following performance measures.

Different classifier results are shown in Table 1

Table 1: Comparison of Performance Measure of Different Classifiers

Techniques	Accuracy (%)	Sensitivity (%)	Specificity (%)
KNN	76.2	77.2	77.5
Neural Network	85.3	84.1	85.3
SVM	86.3	87	87.3
Bayesian	87.3	89.3	89.6

5. Conclusion & Future Work

Proposed system is developed for diagnosing the breast cancer from mammogram images. This system performs this diagnosis in multiple phases. In first phase preprocessing on mammogram image is done to enhance image quality. Then features extraction has been performed. Bayesian classifier has been used for classification. All experiments show that the proposed system gives exceptionally good results. In the future, we will perform to classify the malignancy of breast images.

Acknowledgements

This project was funded by King Saud University, Deanship of Scientific Research, and College of Science Research Center. The authors are thankful to the Faculty of Computer Science and Information Systems at the Universiti Teknologi Malaysia for providing research facilities in conducting this research.

References

[1] M. J. M. Broeders and A. L. M. Verbeek, "Breast cancer epidemiology and risk factors", Q-J-Nucl-Med, 41(3):179–88, **(1997)**.

[2] N. Naveed, T. S. Choi and M. A. Jaffar, "Malignancy and abnormality detection of mammograms using DWT features and ensembling of classifiers", International Journal of the Physical Sciences Vol. 6(8), pp. 2107–2116, **(2011)** April 18.

[3] M. Wallis, M. Walsh and J. Lee, "A review of false negative mammography in a ymptomatic population", Clin. Radiology, vol. 44, pp. 13-15, **(1991)**.

[4] J. Tang, R. Rangayyan, J. Xu, I. El Naqa and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances", Information Technology in Biomedicine, IEEE Transactions on 13(2), **(2009)** pp. 236–251.

[5] G. Kom, A. Tiedeu and M. Kom, "Automated detection of masses in mammograms by local adaptive thresholding", Comput. Biol. Med. 37(1), **(2007)** pp. 37–48.

[6] N. Eltonsy, G. Tourassi and A. Elmaghraby, "A concentric morphology model for the detection of masses in mammography", Medical Imaging, IEEE Transactions on 26(6), **(2007)** pp. 880–889.

[7] N. Naveed, M. A. Jaffar, T. S. Choi, "Segmentation and texture-based classification of breast mammogram images", Microscopy Research and Techniques, DOI: 10.1002/jemt.21070.

[8] X. Wang, B. Zheng, W. F. Good, J. L. King and Y. Chang, "Computer assisted diagnosis of breast cancer using a data-driven bayesian belief network", International Journal of Medical Informatics, 54(2):115–126, **(1999)** May.

[9] S. Timp, "Analysis of Temporal Mammogram Pairs to Detect and Characterise Mass Lesions", PhD in medical sciences, Radboud University Nijmegen, **(2006)**.

[10] M. Mavroforakis, H. Georgiou, N. Dimitropoulos, D. Cavouras and S. Theodoridis, "Significance analysis of qualitative mammographic features, using linear classifiers, neural networks and support vector machines", European Journal of Radiology, 54(1):80–89, **(2004)**.

[11] A. Daskalakis, et al, "An Efficient CLAHE-Based, Spot Adaptive, Image Segmentation Technique for Improving Microarray Genes' Quantification", 2nd International Conference on Experiments/Process/System Modelling/Simulation & Optimization, Athens, **(2007)** July 4-7.

[12] G. Strang, "The Discrete Cosine Transform," SIAM Review, Vol. 41, No. 1, **(1999)**, pp.135-147.

[13] I. T. Jolliffe, "Principal Component Analysis", Springer-Verlag, pp. 487 **(1986)**.

[14] S. M. Stigler, "The history of statistics, Inverse Probability", Harvard University press, **(1986)**, pp. 94-126.

[15] R. O. Duda, P. E. Hart and D.G. Stork, "Pattern classification", 2nd edition, John Wiley and Sons, New York, **(2001)**.

[16] Qurat-ul Ain, S. M. Naqi and M. A. Jaffar, "Bayesian Classification Using DCT Features for Brain Tumor Detection", Lecture Notes in Computer Science, **(2010)**, Volume 6276, Knowledge-Based and Intelligent Information and Engineering Systems, pp. 340-349.

[17] J. Suckling, et al., "The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica", International Congress Series 1069, **(1994)**, pp. 375-378.

[18] R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, M. Roffilli, "A novel featureless approach to mass detection in digital mammograms based on support vector machines", Phys. Med. Biol., 49: 961-975, **(2004)**.

[19] Q. Guo, J. Shao and V. Ruiz, "Investigation of support vector machine for the detection of architectural distortion in mammographic images", J. Phys. Conf. Ser., 15: 88–94, **(2005)**.

[20] P. Miller and S. Astley, "Automated detection of breast asymmetry using anatomical features", in State of the Art in Digital Mammographic Image Analysis. Ser. Machine Perception Artif. Intell., 9: 247-261, **(1994)**.

[21] C. Daniels, G. H. Schaller, N. L. McNeil, P. Salvador, T. Matcalfe, J. S. Caines and N. A. Wadden, "Fundamentals of breast imaging, a multimedia teaching file for residents, technologists, radiologists and physicists", Radiology dept. Dalhousie University, Halifax, Nova Scotia; Radiology dept. University of Toronto, **(1997)**, ISBN: 0-7703-2806-7.

[22] N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms", Phys Med Biol, 43:365–378, **(1998)**.

Authors

Muhammad Talha Naseem Uppal was born in Pakistan. He received the Master degree in Statistics from University of the Punjab, Lahore Pakistan. His PhD in medical imaging from Universiti Teknologi Malaysia (UTM) is in dissertation writing stage.

Currently he is a lecturer in King Saud University, Saudi Arabia. His research interests include medical image processing, classification and image enhancement for medical applications. He is an author or co-author of about 20 multidisciplinary papers published in local and international journals.

Ghazali Bin Sulong is a Professor at the Department of Computer Graphics and Multimedia, Universiti Teknologi Malaysia. His research areas include Image Processing and Pattern Recognition. His current research interests include fingerprint identification, face recognition, image enhancement and restoration, medical image segmentation, digital watermarking and steganograpy.