DCT Based Features for the Detection of Microcalcifications in Digital Mammograms

Ahmed Farag¹, Samia Mashali²

Abstract - In this paper, a set of spectral domain features based on the discrete cosine transform (DCT) of mammograms are extracted from the X-ray image. The extracted features by the proposed methods are exploited to classify regions of interest (ROI’s) into positive ROI’s containing clustered microcalcifications and negative ROI’s containing normal tissues. A three-layer back-propagation neural network is used as a classifier. The results of the neural network for the extracted features are evaluated by using a receiver operation characteristics (ROC) analysis. The proposed technique is shown to be superior to the conventional methods with respect to classification accuracy and computational complexity.

Index Terms – Breast Cancer, clustered microcalcifications, mammography, neural network, Spectral domain image analysis.

I. INTRODUCTION

Breast Cancer is one of the major causes for the increase in mortality among middle-aged women, especially in developed countries[1]. Breast cancer is relatively unusual in individuals less than 35 years of age, but its incidence rises appreciably by age 40 [2]. Although this incidence continues to increase with age, the disease is of particular concern in the 40- to 50-year age group where it is the leading cause of malignancy related deaths in women and nearly the leading cause of death from all medical conditions [3], [4].

Mammography associated with clinical breast examination is the most efficient method for early detection of breast cancer [5]. The radiographic findings in breast cancer are related to the local manifestations of an ever-increasing number of neoplastic cells. These findings fall into two large categories, microcalcifications and masses [6], [7].

Microcalcifications related to malignancy typically arise in ducts that are occluded by neoplasms. Clustered microcalcifications on mammograms are an important sign for early detection of breast cancer.

Although mammography associated with clinical breast examination is the most efficient method for early detection of breast cancer, it is very difficult to interpret X-ray mammograms because of the small difference in image densities of various breast tissues [5]. The diagnosis process using mammograms by radiologists is performed by a visual examination of films for the presence of abnormalities that indicate cancerous changes. Computer aided diagnosis (CAD) can be used as a useful tool for improving the accuracy of the diagnosis process, and for helping the radiologists with film interpretation.

Many methods for computerized detection and classification of microcalcifications employ mathematical representations of the radiographic features exhibited by clustered microcalcifications. Such extractions may be performed at all pixel locations in an image and/or just at those locations previously delineated as suspect in a preprocessing stage. These features are used in the diagnosis phase to help in the diagnosis process.

In computer applications in mammography, features characterizing suspicious image regions are extracted and input to a data classifier (such as an artificial neural network) to predict likely pathology nations of features will, in general, yield different classification performances.

In this paper, a set of features based on spectral domain are used to classify the mammogram. The paper will be organized as follows. Microcalcifications detection techniques will be reviewed in section 2. Section 3 discusses the proposed technique. Results on classification accuracy and computation complexity are discussed in section 4. A conclusion on the proposed work is given in section 5.

II. REVIEW OF MICROCALCIFICATIONS DETECTION TECHNIQUES

The key to high performance in the automatic detection process of microcalcifications lies in using efficient features that help in the detection process. Image processing and classification techniques have been successfully utilized for this purpose. Microcalcifications are tiny granule-like deposits of calcium. The presence of clustered microcalcifications in X-ray mammograms is considered an important indicator for the detection of breast cancer [8].

1 Ahmed Farag is with the Biomedical Engineering Department, Faculty of Engineering Helwan University, Cairo Egypt (e-mail: ahmed.farag@ieee.org).

2 Samia Mashali is with the Computers and Systems Department, Electronics Research Institute, Cairo, Egypt, (e-mail: samia@eri.sci.eg).
Image features can be classified to either as spatial or spectral domain features. Spatial domain features are those features extracted from the spatial characteristics of the image. While spectral domain features, that are proposed in this paper, are extracted based on the spectral characteristics of the region of interest. Classification techniques, which use spatial features, referred to as spatial or textural feature classifications.

The detection of microcalcifications in mammograms has been of great interest to many researchers [5], [9-19]. Chan et al. [9] used difference image, gray level thresholding, and signal extraction techniques to detect microcalcifications in digital mammograms. Wu et al. [10] used a neural network whose input was the mammogram image or an enhanced version of the image to recognize patterns that that might include microcalcifications in digital mammograms. Some other researchers [11] proposed a local area thresholding process for the same problem of detecting microcalcifications. Recently, many researches based on the wavelet transform have been proposed for the detection of microcalcifications [12], [13], [14]. Nishikawa et al have also studied the use of CAD to improve the radiologist’s performance in the diagnosis of clustered microcalcifications [15], [16]. The development of intelligent clinical workstations for CAD of breast cancer have been proposed to show the possibility of using CAD for clinical applications [17], [18].

The spectral features of the image play an important role in image analysis and understanding, with potential applications in medical diagnosis. This paper proposes spectral features based on the discrete cosine transform of the image that can help in the diagnosis process for the detection of microcalcifications.

### III. THE PROPOSED TECHNIQUE

To exploit the spectral features of the digital mammogram image a discrete cosine transform of the image is used to transform the image to the spectral domain to help in the diagnosis process for the detection of microcalcifications.

The mammogram image is segmented into ROI’s. The smoothness or coarseness of an ROI can be interpreted as the distribution of the DCT coefficients on the DCT matrix of the ROI. For example, if the texture is smooth, a pixel and its surrounding pixels will probably have the similar gray values. This means that the higher frequency coefficients of the DCT transform of ROI will be near zero, and the distribution of the coefficients should be concentrated on or near the upper left corner of the DCT matrix. If a texture has fine details, the difference between a pixel and its surrounding pixels will probably be large. This means that the distribution of the DCT coefficients should be spread out along the diagonal of the DCT coefficients matrix. For positive ROI’s containing clustered microcalcifications, the distribution of DCT coefficients tends to spread to the right and/or lower right corner of the DCT coefficients matrix. The DCT coefficients are taken as the spectral features for classification of positive and negative regions of interest. The small MIAS [19] database of mammography images was selected as the database for the proposed work. First step is to extract ROI from each image. The extracted ROI’s are marked to be either containing microcalcifications or normal tissues of the breast.

Each ROI image (either containing normal tissues or microcalcifications) is divided into 8 X 8 blocks of pixels. Each block is first processed by means of the discrete cosine transform (DCT). In so doing, the 64 pixel values are transposed into 64 frequency components. The mathematical basis for discrete cosine transform follows:

\[
S_{uv} = \frac{1}{4} C_u C_v \sum_{x=0}^{7} \sum_{y=0}^{7} S_{xy} \cos \left( \frac{2x+1}{16} \mu \pi \right) \cos \left( \frac{2y+1}{16} v \pi \right)
\]

Where:

\[
C_u, C_v = 1 \quad \text{for } u,v=0
\]

\[
C_u, C_v = \frac{1}{\sqrt{2}} \quad \text{Otherwise.}
\]

Large even areas in an image are reflected as low frequencies, details as high ones. A subset of the coefficients is selected to minimize the computation complexity needed for the classifier and feature extraction processes. To reduce the number of features the Fisher criterion is used [20] to select the most discriminated coefficients.

To evaluate the classification efficiency of the selected features, a three-layer back-propagation neural network was employed as a classifier. A receiver operation characteristics (ROC) analysis was used to evaluate the classification performance of each selected group of features. ROC analysis is based on statistical decision theory and has been applied extensively to the evaluation of clinical diagnosis. The ROC curve represents the relationship between the true- positive fraction (TPF) and the false positive fraction (FPF) for variations of the decision threshold.

The TPF and FPF denote the fraction of the patients actually having the disease in question that are diagnosed as positive and the fraction of the patients actually without the disease in question that are diagnosed as positive, respectively. The area under the ROC curve is used as a measure for the classification system efficiency. The
optimum set of features that result in the best classification performance is selected to be the spectral features that can classify positive and negative microcalcifications in mammograms.

IV. EXPERIMENTAL RESULTS

Forty mammograms were selected from the MIAS database containing twenty normal cases and twenty images with microcalcifications. After calculating the 64 DCT features for each block, the best 5 features are selected using fisher criteria. These features are applied to a three layer back propagation neural network. To draw an ROC curve the network topology was changed to change the classifier threshold. After training of each classifier network, the network was tested with all infected database to calculate the number of true positive (TP) cases and the number of false negative (FN) cases. Then the same network was tested with all normal database to calculate the number of true negative (TN) cases and the number of false positive (FP) cases.

From these data an ROC curve is drawn for this group of features.

The ROC curves for different groups of features are shown in figure (1).

![ROC Curve](image)

Figure 1: ROC curve for different group of DCT features compared to spatial features used in [5].

From the shown curve it is clear that the selected set of spectral features results in higher efficiency for the classifier than conventional methods with respect to classification accuracy.

V. CONCLUSION

A new approach to the problem of microcalcifications detection in digital mammograms by choosing the best group of feature vector based upon the discrete cosine transform is introduced.

The proposed technique is based on using a new criterion, which is based on the minimum number of coefficients to discriminate between the normal and infected cases. Using the proposed criterion with a three layer back propagation neural network classifier, experimental results showed that a higher classification accuracy than ordinary techniques is achieved. The proposed technique also reduces the complexity of the feature extraction process.

REFERENCES


