COMPUTER-AIDED DIAGNOSIS FOR MICROCALCIFICATIONS ANALYSIS IN BREAST MAMMOGRAMS

Dr. Abbas Hanon AL-Asadi
Ahmed Kazim Hamed Al-Saadi
Basrah University, Iraq

Emails: Abbashh2002@yahoo.com, ahmed_kazim2007r@yahoo.com

Abstract

Breast cancer is the main cause of death for women above age 40. In this paper a method is proposed to develop a Computer-Aided Diagnosis (CADx), this method provides a second opinion in microcalcifications diagnostic and making decisions (classify microcalcifications as benign or malignant). The proposed method for microcalcifications diagnosis splits into three steps: The first step extracts the region of interest (ROI). The second step is the features extraction, where we used a set of features from (ROI) by applying wavelet decomposition. The third step is the classification process where discrimination between benign and malignant is performed using a Nearest Neighbor Classifier. The proposed method was evaluated using the Mammographic Image Analysis Society (MIAS) mammographic databases. The proposed method has achieved satisfactory results.

Keywords:

1. INTRODUCTION

Breast cancer is a malignant tumor that starts in the cells of the breast [1]. In modern countries, medical statistics have estimated that one woman on eight will contract a breast cancer. Malign tumors are more likely to appear in breast tissues of women above age 40, and they represent 35% of the abnormalities detected in women breasts. Breast cancer is currently responsible for more than 30% of death by cancer in women, which is about 1% of all deaths worldwide. Men are also concerned by breast cancers since it represents 1.5% of all cancer death in men [2, 3].

The early detection of breast cancer is very important because the treatment of an undeveloped and non-metastasized tumor will not require massive surgical interventions.

Micro-calcifications are deposits of calcium (Ca₃(PO₄)₂OH) in breast tissue [4]. It is small size lesions, typically range in size from 0.05 to 1 mm in diameter. With these dimensions, Micro-calcifications are relatively difficult to detect [1, 5].

There is a high correlation between the presence of Micro-calcifications and breast cancer, particularly when a number of Micro-calcifications grouped together is termed a cluster and it may be a strong indication of cancer. A cluster is defined as at least three Micro-calcifications within a 1 cm² area. Therefore, an accurate detection of Micro-calcifications is essential to any early detection of the majority of breast cancers [6, 7].

In the literature, various numbers of techniques are described to detect and classify the presence of...
Micro-calcifications in digital mammograms as benign or malignant.

Chan et al. [8- 10] investigated a computer-based method for the detection of Micro-calcifications in digital mammograms. The method is based on a difference image technique in which a signal enhanced image is subtracted from a signal from more structured background in the mammogram. Global and local thresholding techniques are then used to extract potential Micro-calcifications signals. Subsequently, signal extraction criteria are imposed on the potential Micro-calcifications to distinguish true positives from noise and artifacts.

Karssemeijer [11- 13] developed a statistical method for detection of Micro-calcifications in digital mammograms. The method is based on the use of statistical models and the general framework of Bayesian image analysis.

Strickland et al. [14-19] developed a method based on undecimated bi-orthogonal wavelet transforms and optimal sub band weighting for detecting and segmenting clustered Micro-calcifications.

Yoshida et al. [20, 21] used decimated wavelet transform and supervised learning for the detection of Micro-calcifications.

Cheng et al. [22] proposed an approach using fuzzy logic for the detection of Micro-calcifications.

Yu and Guan [23] developed a CAD system for the automatic detection of clustered Micro-calcifications through two steps. The first one is to segment potential Micro-calcifications pixels by using wavelet and gray level statistical features and to connect them into potential individual Micro-calcifications objects. The second step is to check these potential objects by using 31 statistical features. Neural network classifiers were used.

Jiang et al. [24] Proposed Genetic Algorithm (GA) technique which is characterized by transforming input images into a feature domain, where each pixel is represented by its mean and standard deviation inside a surrounding window of size 9x9 pixels. In the feature domain, chromosomes are constructed to populate the initial generation and further features are extracted to enable the proposed GA to search for optimized classification and detection of Micro-calcifications clusters via regions of 128x 128 pixels.

2. MAMMOGRAM DATABASE

The Mammography Image Analysis Society (MIAS), which is an organization of UK research groups interested in the understanding of mammograms, has produced a digital mammography database (ftp://peipa.essex.ac.uk).

The database used in experiments of this work was taken from the MIAS because it contains complete information about abnormalities of each mammographic image. The X-ray films in the database have been carefully selected from the United Kingdom National Breast Screening Program and digitized with a Joyce-Label scanning microdensitometer to a resolution of 50 μm x 50 μm. 8bits represent each pixel with 1024x1024 pixel size and at 256 gray levels.

The images are in the grayscale file format (.pgm) (Portable Graymap). The used database contains left and right breast images for 161 patients. Its quantity consists of 322 images, which belong to three types such as Normal, benign and malignant. There are 208 normal, 63 benign and 51 malignant (abnormal) images.

3. WAVELET ANALYSIS

An important branch of CADx methods in mammography employs wavelet transforms for feature enhancement. This work uses the Two-Dimensional Discrete Wavelet Transform (2D-DWT), which can be defined as:

\[ C(a,b) = C(j,k) = \sum_{x \in Z} \sum_{y \in Z} f(x,y)g_{j,k}(x,y) \]  

(1)

With \( a = 2^j \), \( b = k.2^j \), \( j,k \in N \), where \( f \) is the original image, \( g \) is the wavelet function, \( a \) is a scale factor of the wavelet function, \( b \) is a location parameter of the wavelet function, and \( C(a,b) \) is the set of obtained coefficients. Inverse process is calculated by:

\[ f(x,y) = \sum_{x \in Z} \sum_{y \in Z} C(j,k)\psi_{j,k}(x,y) \]
Where $\psi$ is the wavelet function used to reconstruct the image.

The multi-resolution representation carried out by 2D-DWT fragment the frequency spectrum of an image $f(x,y)$ into a low pass subband image $cA^j$ and a set of band-pass sub band images $cDH^j$, $cDV^j$, $cDD^j$, $j = 1, \ldots, L$, where $L$ denote the number of levels for a representation, $cDH^j$ is formed by low pass filtering the rows followed by high pass filtering the columns, and is therefore sensitive to horizontally oriented features. In the same way the $cDV^j$ contains vertically oriented structure, and $cDD^j$ contains primarily diagonal structure [25].

Generally speaking, multi-resolution representations are implemented by a cascade of analysis/synthesis (A/S) filter banks. The discrete wavelet transform uses two different wavelet mothers: $h(x)$ for multi-resolution decomposition (analysis) and $g(x)$ for reconstruction (synthesis) of the original image from its multi-resolution representation. An efficient way to implement discrete wavelet transform using filters was developed by Mallat [26].

Figure (1) shows the implementation of a one-level ($L = 1$) multi-resolution representation of the discrete wavelet transform, which divides orientations into three bands.

As seen in Figure (1), the forward 2D wavelet transform is implemented using a bank of 1D low pass ($h_1(x)$) and high pass ($h_2(x)$) analysis filters. The reconstruction process, or inverse wavelet transform, is likewise computed via 1D synthesis filters, $g_1(x)$ and $g_2(x)$.

**Figure (1): Wavelet decomposition algorithm**

Wavelet-based image decomposition can be interpreted as an image filtering process. For a given image $I$ of size $2^n \times 2^n$, wavelet-based sub band decomposition can be performed as follows: The wavelet filters $h_1(x)$ and $h_2(x)$ are applied to the rows of the image $I$. The filter $h_1(x)$ is a low-pass filter with frequency response $H_1(w)$ and $h_2(x)$ is a high pass filter with frequency response $H_2(w)$. By filtering the image $I$ with $H_1(w)$, low-frequency is obtained information (background).

By filtering the image with $H_2(w)$, the high-frequency information is obtained (edges). After down sampling by a factor of two, two sub bands are obtained: $H_1rI$ and $H_2rI$ (the subscript $r$ suggests that the filters are applied to rows of the image $I$). The filters $H_1(w)$ and $H_2(w)$ are then applied to the columns of the subbands $H_1rI$ and $H_2rI$, followed by down sampling by a factor of two, and the following four sub bands are obtained: $cA^1$, $cDH^1$, $cDV^1$ and $cDD^1$. The sub band $cA^1$ contains the smooth information of the image, and the subbands $cDH^1$, $cDV^1$ and $cDD^1$ contain the detail information of the image. Then the $cA^1$ sub band of the frequency domain is segmented into four subbands at the second level decomposition, and so on.

**4. MICRO-CALCIFICATIONS DIAGNOSIS METHOD**

In this method, the first step is to apply a pre-processing to improve the edge of breast and then segmentation process (Region of interest) for eliminating some regions in the image, which are not useful for the mammographic interpretation.. Then features are extracted after applied DWT and finally classification process is performed for classifying the Micro-calcifications. Flowchart in Figure (2) depicts the steps of this method.
4.1. Preprocessing

This step enhances the breast’s edge before segmenting the region of interest (ROI) from the mammographic images; this process can be performed by using the Logarithm transformations tool for dynamic range manipulation. Logarithm transformations are implemented using the expression:

\[ g = c \times \log(1 + \text{double}(f)) \]

where \( g \) is output image, \( f \) is input image, and \( c \) is a constant.

4.2. Segmentation

The aim of the segmenting of the image is to reduce the data amount consequently, reduce the computational time and the process searching for Regions of Interest that include a lesion with high probability. This step eliminates the parts in the image that are not useful for the mammographic interpretation and extract only the region of the image that corresponds to the breast. We performed this step by binarization the image with automatic threshold for obtaining a binary image and Morphological operations for the extract the region of breast.

4.3. Features Extraction

In an automated analysis system an important goal is to extract features that are able to "summarize" meaningful information in the mammographic image; information that is otherwise distributed among a large number of pixels. Features here are extracted from the ROI based on the wavelet decomposition process. These features are passed to the classification stage. There are three processing steps in the features extraction stage. Features, in our method, are extracted from the coefficients that were produced by the wavelet analysis decomposition. The following steps discuss in details the extracted features.

i. Wavelet Decomposition

In this work, the wavelet decomposition applied on the region of interest. The outputs of wavelet analysis are the decomposition vector \( C \) and corresponding book keeping matrix \( S \). The vector \( C \) consists of horizontal, vertical, and diagonal detail coefficients and one approximation.

ii. Coefficients Extraction

The horizontal, vertical, and diagonal detail and approximation was extracted from the wavelet decomposition structure \([C, S]\). These vectors were extracted at level one of wavelet decomposition.

iii. Energy, Mean, and Standard Deviation Computation

The energy, mean, and standard deviation have been computed for extracted coefficients. The produced values are considered as features (12 features) for the classification process.

4.4. Classification

The classification process is divided into the training phase and the testing phase. In the training phase, labeled data are given. Separately, the data on a candidate region which has already been de-
cided as Micro-calcifications or as normal is given and the classifier is trained. In the testing phase, unknown data are given and the classification is performed using the classifier after training. The number of images which were used in training and testing sets is shown in Table (1). This step can be divided into two processes:

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of image</th>
<th>No. of training sets</th>
<th>No. of testing sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>40</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Abnormal</td>
<td>25</td>
<td>17</td>
<td>8</td>
</tr>
<tr>
<td>Benign</td>
<td>12</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Malignant</td>
<td>13</td>
<td>9</td>
<td>4</td>
</tr>
</tbody>
</table>

1. Features Reduction

After computing the features for every image in training set we reduce the number of features by estimated the mean for each wavelet coefficient at each feature vector. Table (2) and Table (3) are examples of training database that used in classification.

<table>
<thead>
<tr>
<th>Approximation</th>
<th>Horizontal</th>
<th>Vertical</th>
<th>Diagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.9327</td>
<td>0.0094</td>
<td>0.0558</td>
<td>0.002</td>
</tr>
<tr>
<td>99.9728</td>
<td>0.0128</td>
<td>0.013</td>
<td>0.0014</td>
</tr>
<tr>
<td>99.9709</td>
<td>0.0099</td>
<td>0.0178</td>
<td>0.0014</td>
</tr>
<tr>
<td>99.9879</td>
<td>0.0032</td>
<td>0.008</td>
<td>0.00098003</td>
</tr>
<tr>
<td>99.942</td>
<td>0.0138</td>
<td>0.0417</td>
<td>0.0025</td>
</tr>
<tr>
<td>99.9675</td>
<td>0.0132</td>
<td>0.0174</td>
<td>0.0019</td>
</tr>
<tr>
<td>99.9141</td>
<td>0.0091</td>
<td>0.0746</td>
<td>0.0023</td>
</tr>
<tr>
<td>99.965</td>
<td>0.004</td>
<td>0.0302</td>
<td>0.00082275</td>
</tr>
<tr>
<td>99.9782</td>
<td>0.0076</td>
<td>0.013</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

Table (3): Energy Feature for Benign Class

<table>
<thead>
<tr>
<th>Approximation</th>
<th>Horizontal</th>
<th>Vertical</th>
<th>Diagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.9402</td>
<td>0.0075</td>
<td>0.0509</td>
<td>0.0014</td>
</tr>
<tr>
<td>99.9006</td>
<td>0.01</td>
<td>0.0878</td>
<td>0.0016</td>
</tr>
<tr>
<td>99.9674</td>
<td>0.0069</td>
<td>0.0241</td>
<td>0.0016</td>
</tr>
<tr>
<td>99.9398</td>
<td>0.0123</td>
<td>0.0462</td>
<td>0.0016</td>
</tr>
<tr>
<td>99.9299</td>
<td>0.0064</td>
<td>0.0624</td>
<td>0.0013</td>
</tr>
<tr>
<td>99.9649</td>
<td>0.0103</td>
<td>0.0231</td>
<td>0.0018</td>
</tr>
<tr>
<td>99.9305</td>
<td>0.007</td>
<td>0.0606</td>
<td>0.0019</td>
</tr>
<tr>
<td>99.9757</td>
<td>0.0076</td>
<td>0.0151</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

2. Classify

The Nearest Neighbor (NN) classifier has been used to classify classes normal, malignant, and benign. The classifier used the Euclidean distance as a metric between the correspondents normalized wavelet coefficients, as shown in Equation (3).

\[ D_{Euclidian} = \sqrt{\sum_{i=1}^{n} ((x[i] - x[j][i])^2) } \]  

(3)

where \( x \) is the feature vector of unknown pattern, \( x_j \) is the mean vector of class \( j \), and \( n \) is the dimensionality of the feature space. Then the Nearest Neighbor states that the vector \( x \) is to be assigned to the class which has the minimum distance.

5. CADx EVALUATION

This section presents the results achieved in Micro-calcifications diagnosis method. We obtained results of classification among normal, benign and malignant classes using Nearest Neighbor (NN) classifier.

In the beginning, the method of classification is tested twice: once, between normal and Micro-calcifications, and once between benign and malignant. The results were good in classifying the malignant and benign, whereas it was bad in classifying the normal images. Also, this method tested
with one, two, and three levels of the family wavelet (db4) decomposition and was best results with level one.

The classification results of method depending on Table (1) and using the family wavelet (db4) with level one of decomposition were 100% sensitivity (Equation 4), 100% specificity (Equation 5), 100% accuracy (Equation 6), and 100% precision (Equation 7). Table (4) shows the successful rates of classification for normal, benign and malignant classes.

\[
\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (4)
\]

\[
\text{Specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}} \quad (5)
\]

\[
\text{accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{true negatives} + \text{true positives} + \text{false positives} + \text{false negatives}} \quad (6)
\]

\[
\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (7)
\]

Table (4): Successful rates of classification

<table>
<thead>
<tr>
<th>Class</th>
<th>No. of testing sets</th>
<th>No. of successful sets</th>
<th>Successful rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>20</td>
<td>20</td>
<td>100%</td>
</tr>
<tr>
<td>Benign</td>
<td>4</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td>Malignant</td>
<td>4</td>
<td>4</td>
<td>100%</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

In this work, we proposed method for Micro-calcifications diagnosis in mammograms of breast. After implementing this method, we concluded the following:

1. Despite the abundance of theoretical studies in this field, computerized Micro-calcifications detection and diagnosis is one of the things that need a lot of study and practice. This belongs to many reasons. Those come from the great variability in the database mammograms, the use of poor resolution Micro-calcifications mammograms, small number of the available database, some mammograms are still film-based and are read using a light box thus commercial (CADx) systems digitize the film and present markers on a small display or on a separate printout, for the purpose of detection and classification the image is just decomposed to generate features for a classifier and the task of enhancement is more complex as it also requires an image reconstruction.

2. Micro-calcifications are subtle signs of breast cancer and are very difficult to detect and diagnose in the mammographic images because of their small size, low contrast with respect to the normal breast tissue and proximity to the surrounding tissues.

3. Because of the importance of the information in medical images, our goal was to keep this information. Perhaps the uses of wavelet transform characteristics have had a significant impact in achieving this goal.

7. REFERENCES


Engineering, The Faculty of Graduate studies, The University of British Columbia.


