Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural

Rafayah Mousa, Qutaishat Munib*, Abdallah Moussa

Department of Computer Information Systems, University of Jordan, Amman 11942, Jordan

Abstract

The high incidence of breast cancer in women has increased significantly in the recent years. The most familiar breast tumors types are mass and microcalcification. Mammograms—breast X-ray—are considered the most reliable method in early detection of breast cancer. Computer-aided diagnosis system can be very helpful for radiologist in detection and diagnosing abnormalities earlier and faster than traditional screening programs. Several techniques can be used to accomplish this task. In this paper, two techniques are proposed based on wavelet analysis and fuzzy-neural approaches. These techniques are mammography classifier based on globally processed image and mammography classifier based on locally processed image (region of interest). The system is classified normal from abnormal, mass for microcalcification and abnormal severity (benign or malignant). The evaluation of the system is carried out on Mammography Image Analysis Society (MIAS) dataset. The accuracy achieved is satisfied.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Digital mammogram classifier; Breast cancer; Mass tumor; Microcalcification; Wavelet analysis; ANFIS

1. Introduction

The interpretation and analysis of medical images represent an important and exciting part of computer vision and pattern recognition. Developing a computer-aided diagnosis system for cancer diseases, such as breast cancer, to assist physicians in hospitals is becoming of high importance and priority for many researchers and clinical centers. It is a complex process to develop a computer vision system to perform such tasks.

The high incidence of breast cancer in women has increased significantly in the recent years. It is the cause of the most common cancer death in women. It is a leading cause of fatality in women, with approximately 1 in 12 women affected by the disease during their lifetime (Spence, Parra, & Sajda, 2001). In Australia, approximately 1 of 13 women develops the disease (Verma & Zakos, 2000). A report from the National Cancer Institute (NCI) estimates that about one in eight women in the United States (approximately 12.5%) will develop breast cancer during their lifetime (Arun, 2001). Early detection plays a very

important factor in cancer treatment and allows better recovery for most patients. The required medical image for the diagnosing process of breast cancer, mammogram (breast X-ray), is considered the most reliable method in early detection (Arun, 2001; Verma & Zakos, 2000).

Due to the high volume of images to be analyzed by radiologists, and since senior radiologists are rare, reliable radiological diagnosis is not always available and the accuracy rate tends to decrease. A statistics shows that only 20–30% of breast biopsies are proved cancerous (Zaiane, Maria-Luiza, & Alexandru, 2002), and 10% of all cases of breast cancer go undetected by mammography (Bird, Wallace, & Yankaskas, 1992). Moreover, digital mammograms are among the most difficult medical images to be read according to the differences in the types of tissues and their low contrasts. Important visual clues of breast cancer include preliminary signs of masses and microcalcification clusters (Roberts, Kahn, & Haddawy, 1995). Unfortunately, at the early stages of breast cancer, these signs are very subtle and varied in appearance, making diagnosis even difficult to specialists. Therefore, automatic reading of digital medical images becomes highly desirable. It has proven that double reading of the mammogram, by two radiologists, increases the accuracy, but at high costs

---

* Corresponding author. Tel.: +962 7749 8959; fax: +962 6535 5511.
E-mail address: maq@ju.edu.jo (Q. Munib).

(Zaiane, Maria-Luiza, & Alexandru, 2001). Therefore, the motivation of the computer-aided diagnosis systems (Mendez, Tahoces, Lado, & Souto, 1998; Taylor, 1995; Woods, 1994; Yin, Giger, Vyborny, Doi, & Schmidt, 1993) is to assist medical staffs to achieve high efficiency and accuracy.

Radiologists basically look for two types of patterns in mammography: micromicrocalcifications and masses (Arun, 2001). The diagnosis result of tissue is classified into three categories: normal which represents mammogram without any cancerous cell, benign which represents mammogram showing a tumor, but not formed by cancerous cells and malign which represents mammogram showing a tumor with cancerous cells (Verma & Zakos, 2000). It is difficult to distinguish a benign from one that is malignant. Consequently, many unnecessary biopsies are often undertaken due to the high positive false rate (Arun, 2001; Wang & Karayiannis, 1998).

Interpreting medical images that used for diagnosing process involves preprocessing and detection of regions of interest (Arun, 2001). Preprocessing stage deals with image enhancement and noise removal. The enhanced image is then scanned for selected region of interest. Histogram equalization is one of many techniques that used to enhance mammograms. The next stage is to extract features from the region of interest. These features then passed to classifier to decide whether this mammogram normal or abnormal. Many approaches are used to build such classifiers for digital mammograms such as neural network and data mining. The neuro-fuzzy approach is a typical approach for the developing of such types of systems (Mitra & Hayashi, 2000; Nieto & Torres, 2003; Russo & Jain, 2001; Verma & Zakos, 2000). Neural network provides algorithms for learning and classification, whereas fuzzy logic deals with issues reasoning on a higher semantic level.

In this paper, two techniques for building a computer-aided diagnosis system for classification of abnormality in digital mammograms are designed and evaluated. The first one is a neuro-fuzzy classifier based on features extracted from the wavelet analysis of the image. It consists of preprocessing, features extraction and classification stages. Histogram equalization and gray level thresholding techniques are applied for enhancing the images. Features are extracted from the whole image, which represents the unit of classification. We called this technique mammography classifier based on globally processed image. In the classification stage, we apply the Adaptive Neuro-Fuzzy Inference System (ANFIS) (Jang, 1993). The purpose of the system is to classify normal mammogram from abnormal one and to determine abnormal severity in the abnormal one. It could be mass (benign or malign) or microcalcification (benign or malign). In this paper the abnormal cases: mass (circumscribed and speculated and microcalcification) and micocalcification are considered.

Many studies have been made on the problem of breast cancer diagnosing based on digital mammograms (Qian,

Sunden, Sjostrom, Fenger-Krog, & Brodin, 2002; Sameti & Ward, 1996; Verma & Zakos, 2000; Woods, 1994). Verma and Zakos (2000) developed a system based on fuzzy-neural and feature extraction techniques for detecting and diagnosing microcalcifications' patterns in digital mammograms. They investigated and analyzed a number of feature extraction techniques. The following 14 features were used for the proposed method: average histogram, average gray level, energy, modified energy, entropy, modified entropy, number of pixels, standard deviation, modified standard deviation, skew, modified skew, average boundary gray level, difference and contrast. The formula for entropy, energy, skew, and standard deviation were modified so that the iterations started with the first pixel of the pattern and ended at the final pixel and found that a combination of three features, entropy, standard deviation and number of pixels, is the best combination to distinguish a benign microcalcification pattern from a malignant one. The fuzzy technique only detects the center pixel of a microcalcification area. Therefore, it detects other areas that look like a microcalcification. It is up to the user to decide whether the resulting detection is a microcalcififcation or some other area. The back-propagation technique was used for classification of features into benign or malignant.

Zaiane et al. (2001) used neural network and data mining techniques for detection and classification of digital mammograms. Histogram equalization are used to enhance the images. The proposed methods classified the digital mammograms in two categories: normal and abnormal. The data collection they used in their experiments was taken from MIAS (Suckling et al., 1994). The extracted features that used are two existing features (type of the tissue and position of the breast), and four statistical parameters. In their experiments they used 90% of the dataset—322 images—for training the systems and 10% for testing them. The success rate obtained using the neural network (back-propagation algorithm) is 81% on average. On the other hand, it is 69% on average for association rule classifier. In the following research for Zaiane et al. (2002) the data mining classifier is enhanced by applying two pruning methods of rules. They are eliminating the specific rules and keep only those that are general and with high confidence, and prune some rules that could introduce errors at the classification stage. All the extracted features presented in (Zaiane et al., 2002) have been computed over smaller windows of the original image. The classification rate increased to 80%.

Research into the detection of microcalcifications using the wavelet transform has been carried out by McLeod and Parkin (1996). Extraction of possible microcalcifications is firstly achieved by wavelet decomposition of the mammogram using Deubechies wavelets to three levels. This research showed that microcalcifications are mostly prominent in the high-pass subbands of levels 2 and 3, with level 1 containing mostly noise and fine structural detail.

Wang and Karayiannis (1998) proposed an approach for detecting microcalcifications in digital mammograms

employing wavelet-based subband image decomposition. Given that the microcalcifications correspond to high-frequency components of the image spectrum, detection of microcalcifications is achieved by decomposing the mammograms into different frequency subbands, suppressing the low-frequency subband, and, finally, reconstructing the mammogram from the subbands containing only high frequencies. The final images are obtained using subband reconstruction. They used two wavelets from Daubechies' family namely the Daubechies' 4 (db 4) filter and Daubechies's 20 (db 20) filter.

Baguia (2003) proposed a new generalization of the rank nearest neighbor (RNN) rule for multivariate data for diagnosis of breast cancer. The performance of this rule using two well known databases and compare the results with the conventional k-NN rule is studied. The two well-known databases are (i) Wisconsin diagnostics breast cancer (WDBC) database; (ii) Wisconsin breast cancer (WBC) database. They observed that this rule performed remarkably well, and the computational complexity of the proposed k-RNN is much less than the conventional k-NN rules. This approach suffers from a major drawback which is the unavailability of dataset comparable to Wisconsin breast cancer (WBC).

2. Breast cancer

The most familiar tumor types are mass and microcalcification. This section illustrates the features of each type.

2.1. Mass features

Benign and malignant masses are differentiated through mass attributes of margin, density and location (Roberts et al., 1995). Round, low-density masses with smooth, sharply defined margins are considered benign. High-density, stellate, spiculated masses with poorly defined margins are considered malignant. In Fig. 1 we show two examples: (a) mammogram has circumscribed masses, and (b) mammogram has spiculated masses.

2.2. Microcalcification features

The attributes of size, shape, density, distribution pattern, and number of microcalcification are examined when differentiating between benign and malignant microcalcifications (Roberts et al., 1995).

Benign and malignant microcalcifications can occur with or without a mass. Benign microcalcifications are typically large (1–4 mm in diameter), coarse, round or oval, and uniform in size and shape. Their distribution pattern is typically scattered or diffuse. If the microcalcifications are clustered, their number is less than 5 per cluster. Malignant microcalcifications are typically microscopic (<0.5 mm in diameter) and fine, linear branching, stellate-shaped, and varying in size and shape. Their distribution pattern is grouped or clustered, and they are innumerable. The rule of malignancy that when the number of microcalcifications in a cluster is greater (usually more than 5); the likelihood of malignancy become greater (Sickles, 1986). Typically, malignant microcalcifications present with a wide range in size, shape, and density. Mammogram contains micromicrocalcification is shown in Fig. 2.

2.3. Data sources

It is difficult to access real medical images for experimentation due to privacy issue. The data collection that was used in our experiments was taken from the Mammographic Image Analysis Society (MIAS) (Suckling et al., 1994). This same collection has been used in other studies of automatic mammography classification. It consists of 322 images, which belong to three categories: normal, benign and malign, which are considered abnormal.
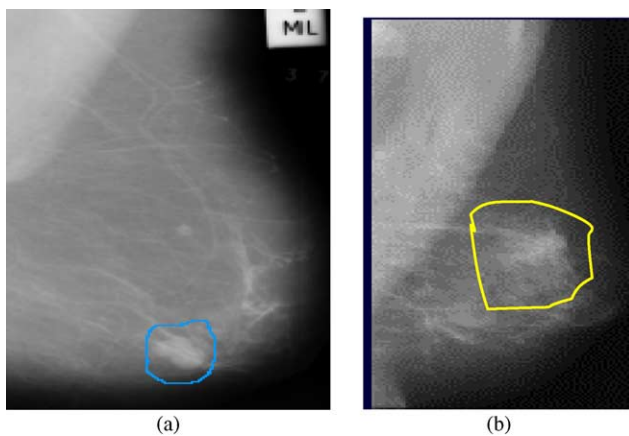
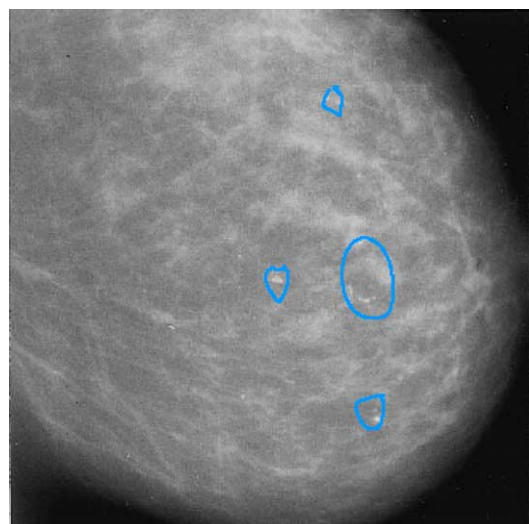Fig. 1. Mammograms contain masses (a) circumscribed masses, and (b) spiculated masses.

Fig. 2. Mammogram contains microcalcifications.

In addition, the abnormal cases are further divided into six categories: circumscribed masses, spiculated masses, microcalcifications, ill-defined masses, architectural distortion and asymmetry. All images are digitized at a resolution of 1024×1024 pixels and eight-bit accuracy (gray level). They also include the locations of any abnormalities that may be present. The existing data in the collection consists of the location of the abnormality (like the center of a circle surrounding the tumor), its radius, breast position (left or right), type of breast tissues (fatty, fatty-glandular and dense) and tumor type if exists (benign or malign).

3. Methodology

The proposed system is built based on wavelet analysis (Portilla & Simoncelli, 2000) of the image and by applying the Adaptive Neuro-Fuzzy Inference System (Jang, 1993; Jang, Sun, & Mizutani, 1997) for building the classifiers. In this section the theoretical background for both approaches are introduced.

3.1. Wavelet analysis

The wavelet transform or wavelet analysis is probably the most recent solution to overcome the shortcomings of the Fourier transform. A wavelet is a waveform of effectively limited duration that has an average value of zero (Kaiser, 1994). Wavelet analysis is the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet. The use of a fully scalable modulated window solves the signal-cutting problem. The spectrum is calculated for the window in each time it shifted. Then this process is repeated many times with a slightly shorter (or longer) window for every new cycle (Mallat, 1989). Wavelet analysis is based on three properties: orthogonal, quadratic filter and filter bank.

3.1.1. Orthogonal wavelets

Two functions f and g are said to be orthogonal to each other if their inner product is zero as show in Eq. (1). The symbol * mean a convolution operation

$$\langle f(t), g(t) \rangle = \int_a^b f(t) g^*(t) \mathrm{d}t = 0 \tag{1}$$

3.1.2. Quadrature mirror filter (QMF)

If a signal $f(x)$ entered to the QMF, the output will be two parts: the first is the output of the low-pass filter while the second is the output of the high-pass filter. Both outputs meet the orthogonal property which means that the output of the high-pass filter cannot be seen in the output of the low-pass filter.

3.1.3. A band-pass filter

The translations of the wavelets are of course limited by the duration of the signal under investigation so that we have an upper boundary for the wavelets. Two issues appear related to dilation: how many scales to analyze our signal are needed, and how to get a lower bound. The answer of this question is by looking at the wavelet transform as band-pass like spectrum. This means that a time compression of the wavelet by a factor of 2 will stretch the frequency spectrum of the wavelet by a factor of 2 and also shift all frequency components up by a factor of 2.

3.1.4. The scaling function

The wavelet function ψ is determined by the high-pass filter, which also produces the details of the wavelet decomposition. There is an additional function associated with the wavelet called scaling function, φ. The scaling function is determined by the low-pass spectrum, and thus is associated with the approximations of the wavelet decomposition. In the same way that iteratively upsampling and convolving the high-pass filter produces a shape approximating, the wavelet function, iteratively upsampling and convolving the low-pass filter produces a shape approximating the scaling function (Kaiser, 1994). Consider the set of expansion functions composed of integer translations and binary scaling of the function, $\varphi(x)$; that is the set $\{\varphi_{j,k}(x)\}$ where:

$$\varphi_{j,k}(x) = 2^{2/j} \varphi(2^j x - k) \tag{2}$$

Here, k determines the position of $\varphi_{j,k}(x)$ along the x-axis and $2^{2/j}$ controls its amplitude. Because the shape of $\varphi_{j,k}(x)$ changes with j, $\varphi(x)$ is called a scaling function.

3.1.5. Wavelet transforms in two dimensions

In 2D, a two dimension scaling functions, $\varphi(x,y)$, and three 2D wavelets, $\psi^H(x,y)$, $\psi^V(x,y)$ and $\psi^D(x,y)$ are required, as show in the following Eqs. (3)–(6):

$$\varphi(x, y) = \varphi(x)\varphi(y) \tag{3}$$

$$\psi^H(x, y) = \psi(x)\varphi(y) \tag{4}$$

$$\psi^V(x, y) = \psi(y)\varphi(x) \tag{5}$$

$$\psi^D(x, y) = \psi(x)\psi(y) \tag{6}$$

These wavelets measure functional variations—intensity or gray level variations for images—along different directions: ψ^H measures variations along columns (for example, horizontal edges), ψ^V responds to variations along rows (like vertical edges), and ψ^{DV} corresponds to variations along diagonals.

Given the separable 2D scaling and wavelet functions, extension of the 1D DWT to two dimensions is straightforward. We first define the scaled and translated basis functions

$$\varphi_{j,m,n}(x, y) = 2^{j/2} \varphi(2^j x - m, 2^j y - n) \tag{7}$$

$$\psi^i_{j,m,n}(x,y) = 2^{j/2}\psi^i(2^j x - m, 2^j y - n) \qquad (8)$$

where index i identifies the directional wavelets in Eqs. (4)–(6). Rather than an exponent, i is a superscript that assumes the values H, V, and D. The discrete wavelet transform of function $f(x,y)$ of size $M \times N$ is then

$$w_\varphi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{x=0}^{N-1} f(x,y)\varphi_{0,m,n}(x,y) \qquad (9)$$

$$w^i_\psi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{x=0}^{N-1} f(x,y)\psi^i_\psi(x,y) \qquad (10)$$

As in the 1D case, j_0 is an arbitrary staring scale and the $w_\varphi(j_0,m,n)$ coefficients define an approximation of $f(x,y)$ at scale and the j_0. The $w^i_\psi(j,m,n)$ coefficients add horizontal, vertical and diagonal details for scales $j \geq j_0$. We normally let $j_0=0$ and select $N=M=2^j$ so that $j=0,1,2\ldots,J-1$ and $m,n=0,1,2,\ldots 2^j-1$.

The decomposition of approximation coefficients vectors are obtained by convolving signal with the low-pass filter Lo_D for approximation, and with the high-pass filter Hi_D for detail, followed by dyadic decimation.

3.2. Adaptive neuro-fuzzy inference system—ANFIS

This system implements a first-order Sugeno-like fuzzy inference system in a five-layer network structure (Jang, Sun, & Mizutani, 1997). Back-propagation is used to learn the antecedent membership functions, while least mean squares algorithm determines the coefficients of the linear combinations in the consequent of the rule. Here the min and max functions in the fuzzy system are replaced by differentiable functions. The rule base must be known in advance, as ANFIS adjusts only the membership functions of the antecedent and consequent parameters. ANFIS can be easily implemented by flexible neural network simulators, and hence is attractive for application purposes. The structure of ANFIS ensures that each linguistic term is represented by only one fuzzy set.

3.2.1. ANFIS structure

Consider a Sugeno type of fuzzy system having the rule base

1. If x is A_1 and y is B_1, then $f_1 = p_1 x + q_1 y + r_1$
2. If x is A_2 and y is B_2, then $f_2 = p_2 x + q_2 y + r_2$

Here, p, q and r is the parameter set of the rule. Let the degree membership functions of fuzzy sets A_i, B_i, $i=1,2$, be μ_{Ai}, μ_{Bi}.

In evaluating the rules, choose *product* for T-norm (logical *and*).

1. Evaluating the rule premises results in this formula where w_i is the firing streanght of the rule.

$$i = 1,2; \quad W_i = \mu_{Ai}\mu_{Bi}(y) \qquad (11)$$

2. Evaluating the implication and the rule consequences gives

$$f(x,y) = \frac{w_1(x,y)f_1(x,y) + w_2(x,y)f_2(x,y)}{w_1(x,y) + w_2(x,y)} \qquad (12)$$

The function without arguments

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2} \qquad (13)$$

The ratio of the ith rule's firing strength to the sum of all rules' firing strengths is computed by the formula:

$$\bar{w}_i = \frac{w_i}{w_1 + w_2} \qquad (14)$$

Then can be written as:

$$f = \bar{w}_1 f_1 + \bar{w}_2 f_2 \qquad (15)$$

Equivalent ANFIS architecture for a two-input first-order Sugeno fuzzy model with two rules is shown in Fig. 3. The output of the ith node in layer 1 is denoted as $O_{1,i}$.

Layer 1. Every node i in this layer is an adaptive node with a node function where x (or y) is the input to node I and
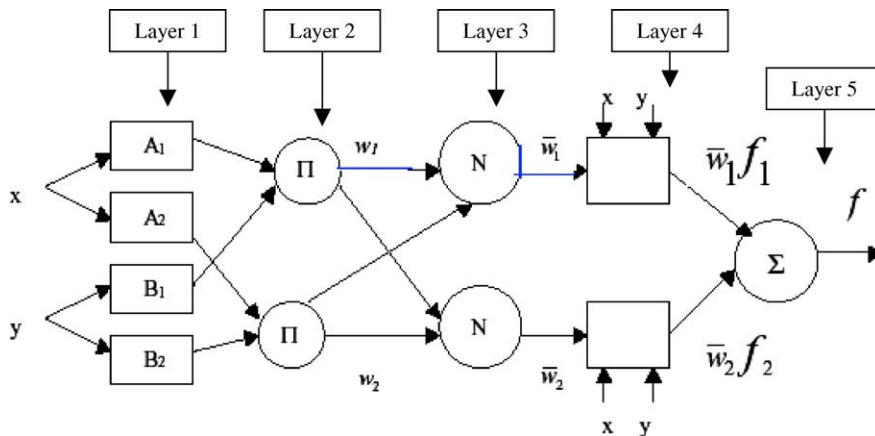

Fig. 3. Equivalent ANFIS architecture for a two-input first-order Sugeno fuzzy model with two rules.

A_i (or B_i) is a linguistic label associated with this node

$$O_{1,i} = \mu_{Ai}(x), \text{ for } i = 1, 2 \quad \text{or}$$
$$O_{1,i} = \mu_{Bi-2}(x), \text{ for } i = 3, 4 \quad (16)$$

Layer 2. Every node in this layer is a fixed node labeled II, whose output is the product of all the incoming signals:

$$O_{2,i} = w_i = \mu_{Ai}(x)\mu_{Bi}(x), \quad i = 1, 2 \quad (17)$$

Each node output represents the firing strength of a rule.

Layer 3. Every node in this layer is a fixed node labeled N. The ith node calculates the ratio of the ith rule's firing strength to the sum of all rules' firing strengths:

$$O_{3,i} = \bar{w}i = w_i/(w_1 + w_2), \quad i = 1, 2 \quad (18)$$

Layer 4. Every node i in this layer is an adaptive node with a node function

$$O_{4,i} = \bar{w}if_i = \bar{w}i(p_i x + q_i y + r_i) \quad (19)$$

where $\bar{w}i$ is a normalized firing strength from layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set of this node.

Layer 5. The signal node in this layer is a fixed node labeled Σ which computes the over all output as the summation of all incoming signals

$$\text{Overall output, } O_{5,1} = \sum_i \bar{w}if_i = \frac{\sum_i w_if_i}{\sum_i w_i} \quad (20)$$

4. The proposed system

The proposed system consists from three stages: preprocessing, feature extraction and classification process. Three techniques are used to enhance the mammograms: image pruning, histogram equalization and global gray level thresholding. Also, feature extraction consists of five steps: image decomposition, coefficient extraction, normalization, energy computation and coefficients reduction. In this section, the proposed system prototype and block diagram are introduced.

4.1. Preprocessing stage

Mammograms are images difficult to interpret, and a preprocessing phase of images is necessary to improve the quality of the images and make the feature extraction phase more reliable. This section introduces the preprocessing techniques before the feature extraction stage. The pre-processing stage consists of two main phases, which are used together. The first phase involves the removal of background information and unwanted parts from the image, while the second phase deals with enhancing the contrast of suspicious areas in the image.

4.1.1. Image pruning

In the MIAS dataset, we had images that were very large (has size 1024×1024) and almost 50% of the whole image
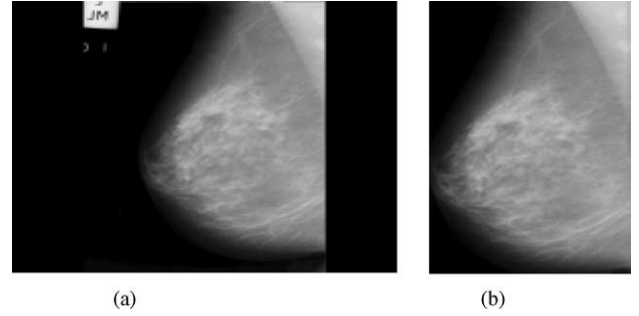

Fig. 4. Preprocessing phase on an example image: (a) original image (b) crop operation.

comprised of the background with a lot of noise. In this phase, we applied a cropping operation to the image to prune the images with the help of the crop operation in Image Processing. Cropping cuts off the unwanted portions of the image. Thus, almost all the background information and most of the noise are eliminated. An example of cropping that eliminates the label on the image and the black background is given in Fig. 4(a,b). The new size of the cropped images is (800×800).

4.1.2. Global gray level thresholding

With global thresholding, the pixels between a pre-selected upper-threshold and lower-threshold of the gray level histogram are retained and all others are set zero (Gonzalez & Woods, 2002). To apply this technique, upper and lower thresholds are determined according to be sure that the region of interest pixels values are between these thresholds. It returns the color values of specified image pixels. In our case, we select the value (240) as upper-threshold and (120) as lower-threshold. The effect of applying this technique is shown in Fig. 5.

4.1.3. Histogram equalization

Histogram equalization method is a well-known gray scale manipulation technique. In histogram equalization, the goal is to map the input image to the output image so that
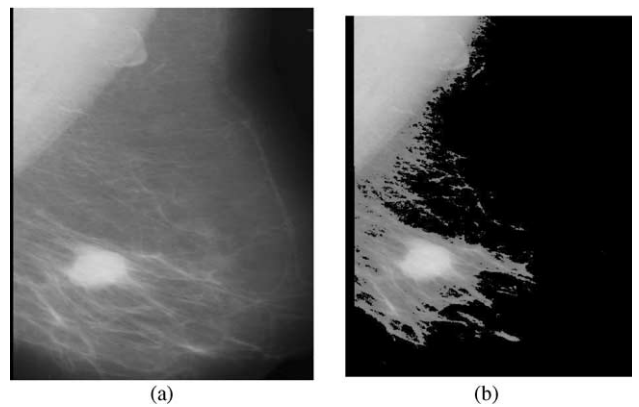

Fig. 5. Effect of the gray level thresholding of the mammogram (a) original image (b) enhanced image.
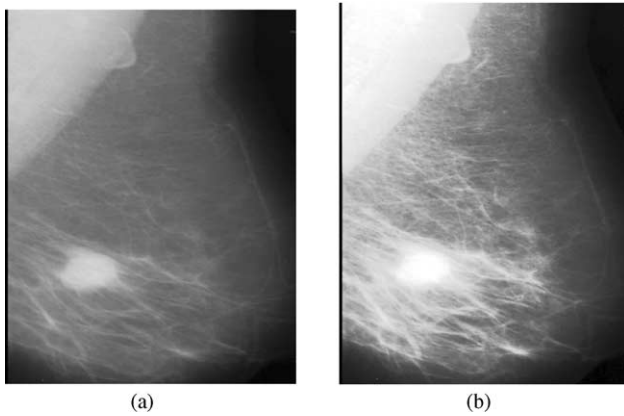
Fig. 6. Effect of histogram equalization: (a) original image and its histogram (b) enhanced image and its histogram.

gray values in the output image are uniformly distributed. For most practical images, gray values need to be redistributed. In histogram equalization we try to spread gray values uniformly over the full gray-scale range. It increases the contrast range in an image by increasing the dynamic range of gray levels (Gonzalez, 2002). Fig. 6 shows the effect of the histogram equalization.

4.2. Features extraction

Features are extracted from the enhanced images based on the wavelet decomposition process. These features are passed to the classification stage. There are five processing steps in the features extraction stage. Features, in our system, are extracted from the coefficients that were produced by the wavelet analysis decomposition. In this section we discuss these steps.

4.2.1. Wavelet decomposition

In the first step, coefficients vector are extracted from the wavelet decomposition of the image. The decomposition operation returns the wavelet decomposition of the image at predefined scale, using the wavelet name, as Daubechies. Outputs are the decomposition vector C and the corresponding bookkeeping matrix S. The decompsostion vector consists from three detail coefficients vector, horizontal detail coefficients, vertical detail coefficients and diagonal detail coefficients, and one approximation are row vectors.

In this paper, the wavelet 'db 4' is used. The function *wmaxlev*, in Matlab Wavelet toolbox, is applied to determine the maximum useful wavelet decomposition scale (N) that the image decomposed at this scale. It helps to avoid unreasonable maximum scale values.

It returns the maximum scale decomposition of image of size X using the wavelet named in the string *wname*. The maximum scale decomposition of image is determined according to number of scales that contain irredundant information. In this thesis, we had images that were of size (800×800).

4.2.2. Coefficients extraction

In this step, we extracted horizontal, diagonal, and vertical details coefficients from the wavelet decomposition structure [C,S]. It returns the horizontal H, vertical V, and diagonal D detail coefficients vectors at scale N. These vectors are extracted at each scale without scale one. We ignore scale 1 coefficients because it contains high frequency details and noise. These details are insignificant information that will not affect the classification accuracy and image quality. We compute the image quality after zeroing coefficients in scale one and it was 99% of the original image.

4.2.3. Normalization

In the third step, the coefficients vectors (H, V and D) for scales two to five are normalized after extracted. The normalization process is achieved by dividing each vector by its maximum value. The results of this operation is that all vectors values become less than or equal one. The normalization process is used to simplify the coefficients value.

4.2.4. Energy computation

We compute the energy for each vector by squaring every element in the vector. The produced values are considered as features for the classification process.

4.2.5. Features reduction

Since the image has a large size, it produces high number of coefficients. Therefore, at the last phase, we reduce the number of features by summing a predefined number of energy values together. In the proposed techniques, we use three approaches:

- low number of features by summing (1000) energy values per feature;
- high number of feature by summing (100) energy values per feature.

While in the locally processed image technique (35) energy values per feature is used. The processing steps are depicted in Algorithm 3.3. These features are used to build the mammogram classifier. The sequence of processing steps is shown in Fig. 7.

5. System structure

In this paper, three techniques for building a computer-aided diagnosis system for classification of abnormality in digital mammograms are proposed. The first is a neuro-fuzzy classifier based on features extracted from the wavelet analysis of the whole image. It is called globally processed image. In the second, the region of interest is cropped from the image then features are extracted from this region and passed to the classifier. Therefore, we called this technique
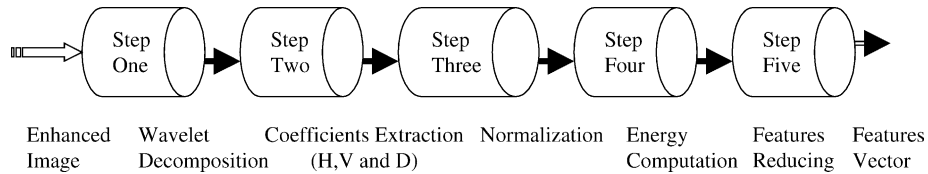
Fig. 7. Feature extraction stage.

the locally processed image. In Section 5.1, the prototype and block diagram of the techniques are explained.

5.1. The neuro-fuzzy classifier based on features extraction

In the classification stage of the globally and locally processed image, we apply the ANFIS technique. The mammography systems block diagram is shown in Fig. 8.

We developed three approaches for the classification stage that applied on the globally and locally processed image techniques. These approaches and their block diagrams are discussed in the next sections. In the classification stage, ANFIS classifier in every phase is trained at specific number of training set in each category.

5.2. Globally processed image vs locally processed image

In the globally processed image technique, we consider the mammogram size, as it is provided in the MIAS dataset, as the input to the system. On the other hand, in the locally

processed image technique, the region of interest is cropped from the image and considered as input to the system. The first approach is analogues to radiographers' mission who investigate the whole image. Fig. 9 shows the difference between the two approaches.

5.3. Classification stage

In this stage, we build a classifier with three phases. In the first one, the ANFIS classifier is applied to classify mammograms into normal and abnormal categories. The mammogram is considered abnormal if it contains tumor (mass or microcalcification). If the result for evaluating the tested mammogram is abnormal; it is entered to the next classification stage to determine if it contains mass or microcalcification tumor. Finally, the abnormal mammogram is classified into malignant or benign in the third stage. The result in each classifier is computed by evaluating the tested mammogram features and then computing the
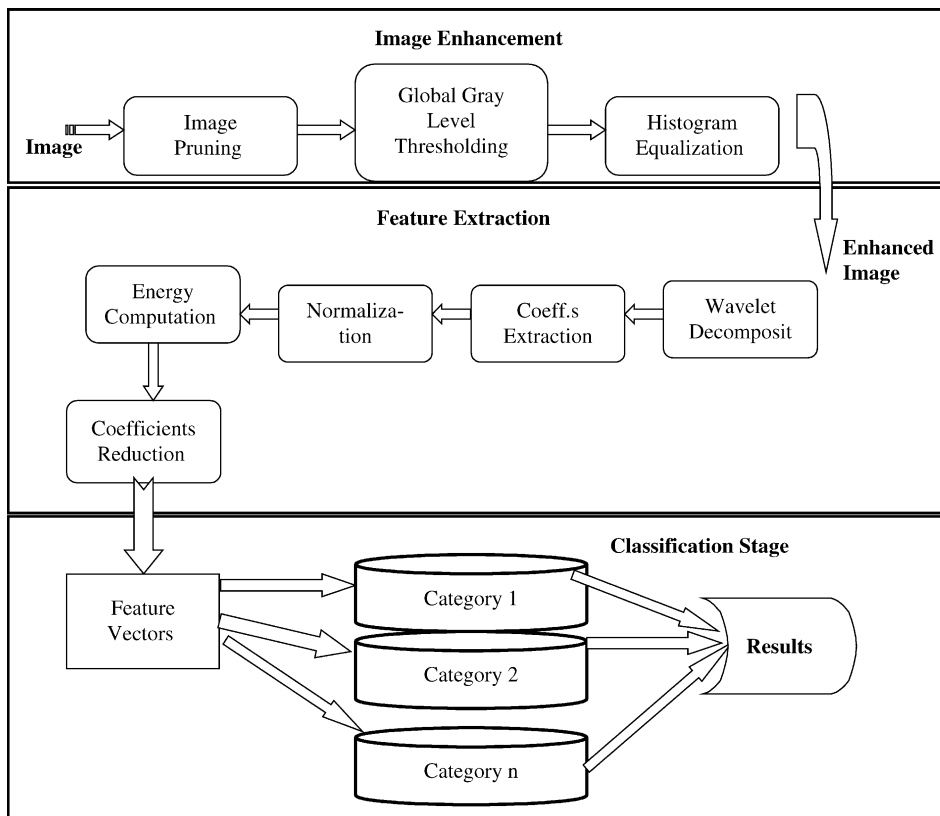
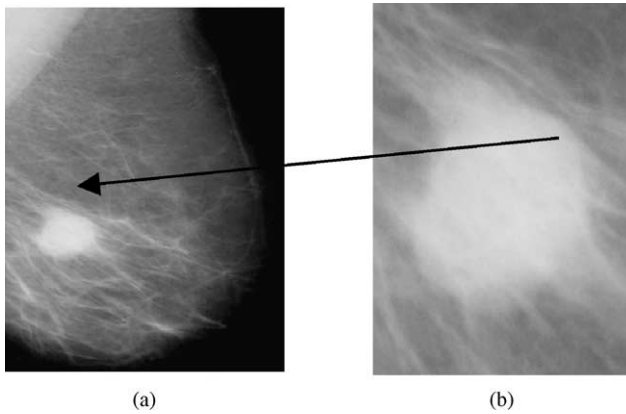

Fig. 8. Digital mammography system.

Fig. 9. Mammogram with mass (a) the unit of classification globally processed image technique (b) the unit of classification in the locally processed image technique (region of interest).

minimum error between the tested one and the output results to each classifier. Fig. 10 explains this approach.

6. Experimentation results

In this section, many experiments are run based on the three approaches that were introduced in the previous chapter. Features are from levels (2–5) and different combinations of these levels are investigated in all simulations. The summation of 100 and 1000 coefficients per feature are tested. The success average is computed, in each experiment, by dividing the number of right classified images at the number of all tested images. Each classifier is trained at different number of iterations. The generalize bell membership function 'gbellmf' and two membership rules are used because they show better results compared with other parameters. Furthermore, the best results achieved at number of iterations (1000 and 5000). All experiments are set with the same neuro-fuzzy parameter settings so there is consistency across all experiments.

6.1. Hardware and software

The system was implemented in MatLab version 6.5. The recognition training and tests were run on a modern standard PC (1.8 GHz AMD processor, 128 MB of RAM) running under Windows 2000.
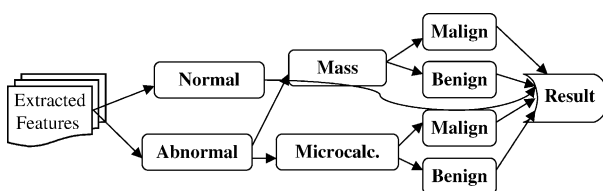

Fig. 10. Classification stage block diagram.

Table 1
Number of training and testing set

Category	No. of training set		Category	No. of training set	
Normal	40	100	Abnormal	40	24
Mass	20	20	Microcalcification	18	6
Mass-benign	10	14	Mass-malignant	10	6
Microcalcification benign	8	4	Microcalcification malignant	8	4

6.2. Globally processed image

The simulation that are run based on 1000 coefficients per feature has achieved poor results because of summing 1000 coefficients per feature, the variations between abnormal and normal coefficients are hidden and become difficult to distinguish. In the next simulation, experiments are run based on the summation of (100) coefficients per feature. The numbers of training and testing sets are shown in Table 1. Table 2(a) shows the classification rate for normal and abnormal categories while the simulation results for mass and microcalcification categories are shown in Table 2(b). The simulation results for tumor severities (benign and malign) are shown in Table 2(c). Results are improved in these experiments compared to the previous approaches.

6.3. Locally processed image technique

The summation of 35 coefficients per feature is considered. The same experimental parameters used in the globally processed image approach are used in this approach. Results show better performance than the first approach. The numbers of training and testing sets are shown in Table 3. Table 4(a) this shows the classification rate for normal and abnormal categories. The simulation

Table 2(a)
Classification rates for normal and abnormal categories

Scales	Stage 1—classifcation rate		
	Normal (%)	Abnormal (%)	Average (%)
2–4	73	75	73.4
3–4	82	79.2	81.4
3	54	33	50

Table 2(b)
Classification rates for mass and microcalcification categories

Scales	Stage 2—classifcation rate		
	Mass (%)	Microcalcifica-tion (%)	Average (%)
2(d)–4	50	100	61
3–4	85	66.6	80.7
2–3	70	83.3	73

Table 2(c)
Classification rates for abnormal severities

Scales	Stage 2—abnormal severity					
	Mass			Microcalcification		
	Benign (%)	Malig-nant (%)	Average (%)	Benign (%)	Malig-nant (%)	Average (%)
2–3	64.3	83.4	70	50	100	75
2–4	71.4	83.4	75	50	75	62.5
2–5	85.7	66.7	80	25	75	50

results for mass and microcalcification categories are shown in Table 4(b). Finally, the experiment results for classifying tumor severities (benign and malignant) are shown in Table 4(c).

6.4. Performance comparison

The first approach tested is neuro-fuzzy classifiers with the summation of 1000 coefficients per feature. This approach does not achieve high performance. Then, the summation of coefficients per feature is reduced to 100. The classification rates are improved by this increasing of features. Finally, the region of interest is extracted from each mammogram and the summation of 35 coefficients per feature is considered. Best results are achieved by building classifiers based on DSSC approach architecture. This approach achieves very well classification rate. Fig. 11 shows the average results of the neuro-fuzzy techniques and approaches.

7. Conclusion

In this paper we have presented and discussed two techniques for building a computer-aided diagnosing system for classification of abnormality in digital mammograms. We have investigated and analyzed wavelet transform for image enhancement and features extraction, and the ANFIS algorithm for classification process. This research has shown that this method is very effective for the automatic detection and classification of abnormalities in digital mammogram. We have examined and compared several algorithms for each technique. The evaluation of the system is carried out on MIAS dataset. Our systems achieved very promising results. Mammography is one of the best methods in breast cancer detection, but in some cases, radiologists

Table 3
Number of training and testing set

Category	No. of training set		Category	No. of training set	
Normal	90	50	Abnormal	42	22
Mass	24	16	Microcalcification	18	6
Mass-Benign	14	10	Mass-Malignant	10	6
Microcalcification Benign	8	4	Microcalcification Malignant	8	4

Table 4(a)
Classification rates for normal and abnormal categories

Scales	Stage 1—classifcation rate		
	Normal (%)	Abnormal (%)	Average (%)
2–3	72	54.5	66.6
2–4	80	72.7	77.7
3–4	76	68.2	73.6

Table 4(b)
Classification rates for mass and microcalcification category

Scales	Stage 2—classifcation rate		
	Mass (%)	Microcalcification (%)	Average (%)
2–3	87.5	83.3	85.4
2–4	71.4	66.7	64.6
2–5	81.25	66.7	77

Table 4(c)
Classification rates for mass and microcalcification severities

Scales	Stage 3—abnormal severity					
	Mass			Microcalcification		
	Benign (%)	Malig-nant (%)	Average (%)	Benign (%)	Malig-nant (%)	Average (%)
2–3	70	66.7	68.7	75	100	87.5
2–4	80	83.3	81.2	75	25	50
3–4	100	83.3	93.7	50	25	37.5

cannot detect tumors despite their experience. Such computer-aided methods like those presented in this paper could assist medical stuff and improve the accuracy of detection.

The classification rate of microcalcification cases achieves the best performance with features extracted from levels 2–3 because microcalcification is small and represented as high frequency information—details—which embodied in the highest levels by wavelet decomposition. On the other hand, the classification rate of mass cases achieve the best performance with features extracted from levels 3–4 because masses have larger sizes and more clear, and it is represented by low frequency information which embodied in the lowest levels by wavelet decomposition.
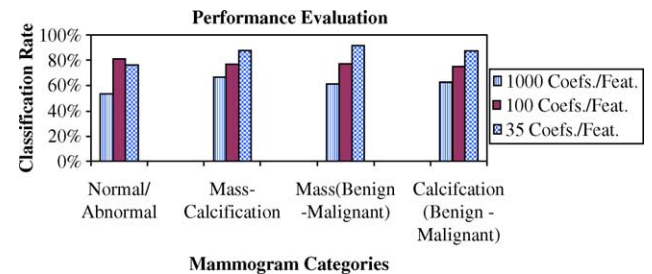

Fig. 11. The average results of the globally processed image technique with 1000 and 100 coefficients per feature, and locally processed image technique.

References

Arun, K. (2001). *Computer vision fuzzy-neural systems*. Englewood Cliffs, NJ: Prentice-Hall.

Baguia, C. (2003). Breast cancer detection using rank-nearest neighbor classification rules. *Pattern Recognition*, 36, 367–381.

Bird, R., Wallace, T., & Yankaskas, B. (1992). Analysis of cancers missed at screening mammography. *Radiology*, 178, 234–247.

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Jang, R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on System, Man and Cybernetics*, 23(6), 181–198.

Jang, R., Sun, C., & Mizutani, E. (1997). *Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence*. Englewood Cliffs, NJ: Prentice-Hall.

Kaiser, G. (1994). *A friendly guide to wavelets*. Boston: Birkhäuser.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(28(4)), 357–381.

McLeod, G., & Parkin, G. (1996). Automatic detection of clustered microcalcifications using wavelet. *The third international workshop on digital mammography, Chicago*.

Mendez, A., Tahoces, P., Lado, M., & Souto, J. (1998). Computer-aided diagnosis: Automatic detection of malignant masses in digitised mammograms. *Medical Physics*, 25, 109–131.

Mitra, S., & Hayashi, Y. (2000). Neuro-fuzzy rule generation: Survey in soft computing framework. *IEEE Transaction on Neural Networks*, 11(3), 748–757.

Nieto, J., & Torres, A. (2003). Midpoint for fuzzy sets and their application in medicine. *Artificial Intelligence in Medicine*, 27, 321–355.

Portilla, J., & Simoncelli, E. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Computer Vision*, 40(1), 49–71.

Qian, W., Sunden, P., Sjostrom, H., Fenger-Krog, H., & Brodin, U. (2002). Comparison of image quality for different digital mammogram image processing parameter settings versus analogue film. *Electromedica*, 71(1), 57–73.

Roberts, M., Kahn E., & Haddawy, P. (1995). Development of a Bayesian network for diagnosis of breast cancer. *IJCAI-95 workshop on building probabilistic networks*.

Russo, M., & Jain, L. (2001). *Fuzzy learning and application*. Englewood Cliffs, NJ: Prentice-Hall.

Sameti, M., & Ward, R. (1996). A fuzzy segmentation algorithm for mammogram, partitioning. *Third international workshop on digital mammography, Chicago*.

Sickles, E. A. (1986). Breast calcifications: Mammographic evaluation. *Radiology*, 160, 289–293.

Spence, D., Parra, L., & Sajda, P. (2001). Detection, synthesis and compression in mammographic image analysis using a hierarchical image probability model. *Artificial Intelligence in Medicine*, 25(31), 365–371.

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., et al. (1994). The mammographic images analysis society digital mammogram database. *Exerpta Medical International Congress Series*, 1069, 375–378.

Taylor, P. (1995). *Decision support for image interpretation: A mammography workstation Information processing in medical imaging*. Dordrecht: Kluwer Academic Publishers.

Verma, K., & Zakos, J. (2000). A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. *IEEE Transactions on Information Technology in Biomedicine*, 16, 219–223.

Wang, T., & Karayiannis, N. (1998). Detection of microcalcification in digital mammograms using wavelets. *IEEE Transaction Medical Imaging*, 51(38(4)), 112–116.

Woods, S. K. (1994). *Automated image analysis techniques for digital mammography*. PhD Thesis. University of South Florida.

Yin, F., Giger, M., Vyborny, C., Doi, K., & Schmidt, R. (1993). Comparison of bilateral-subtraction and single-imageprocessing techniques in the computerised detection of mammographic masses. *Investigative Radiology*, 28(142), 179–183.

Zaiane, O., Maria-Luiza, A., & Alexandru, C. (2001). Application of data mining techniques for medical image classification. In *Proceedings of second international workshop on multimedia data mining (MDM/KDD') in conjunction with seventh ACM SIGKDDS, USA*.

Zaiane, O., Maria-Luiza, A., & Alexandru, C. (2002). Mammography classification by an association rule-based classifier. In *Proceedings of second international workshop on multimedia data mining (MDM/KDD') in conjunction with seventh ACM SIGKDD, USA*.