Computer Aided Diagnosis in Digital Mammography

Akram I. Omara, Ahmed S. Mohamed, Abo-Bakr M. Youssef, and Yasser M. Kadah
Biomedical Engineering Department, Cairo University, Giza, Egypt
e-mail: akramamara@k-space.org

Abstract- **Mammogram – breast x-ray is considered the most effective, low cost, and reliable method in early detection of breast cancer. Although general rules for the differentiation between benign and malignant breast lesion exist, only 15 to 30% of masses referred for surgical biopsy are actually malignant. We are introducing, as an aid to radiologists, a computer diagnosis system, which could be helpful in diagnosing abnormalities faster than traditional screening program without the drawback attribute to human factors.**

The techniques used -in this paper-for feature extraction is based on the wavelet decomposition of locally processed image (region of interest). Both the wavelet coefficients and the statistical measures of different wavelet detail levels are used as features that describe effectively any normal and abnormal region. Two Techniques were used for the classification stage The minimum distance classifier and the voting K-Nearest Neighbor classifier .

Keywords - mammogram; Breast cancer; classifier; Wavelet analysis

I. INTRODUCTION

Breast cancer is a leading cause of fatality among all cancers for women in the 35 to 55 age group. The expected rate is increasing in many countries especially in the United State, where it is estimated that cancer affects three out of four families. There is no known way of preventing breast cancer but early detection allows treatment before its spread to other parts of the body. There are several ways for detecting and diagnosing breast cancer such as self-examination and clinical exams, mammography, and surgical biopsy. Mammography is considered to be safe, less harmful compared to biopsy, and more accurate than self examination where the tumor can not be detected before it can be felt. Mammography is then the best method for early detection of breast cancer, and the percentage of patient that can be cured at early stage is usually high- mortality could be decreased by as much as one third if all women in the appropriate age groups were regularly screen .

In the screening programs, a large number of mammograms must be red. Although the criteria for malignancy are reasonably well established, the application of such criteria is often quite subjective and increases the burdeon on each physician, so abnormality may be over looked due to fatigue.

Moreover, proper evaluation is a time consuming task for the radiologist due to both the large number of mammograms to be red and huge amount of information embedded within, and usually required a review of current and prior films (if available) by a magnifying glass. The mammograms are extensively searched for signs of abnormalities but these signs are very subtle and varied in appearance making diagnosis difficult even to specialist [1], which is the main cause of many missed diagnosis that can be mainly attribute to human factors such as subjective or varying decision criteria, distraction by other image feature or simple oversight.

Since senior radiologists are rare and mammogram alone cant prove that a suspicious area is tumorous , malignant or benign, and since digital mammograms are among the most difficult medical image to be read according to the differences in the type of tissues and their low contrast, the surgical biopsy is applied for closer examination.

Studies indicate that approximately 10 to 30% of breast cancer cases are missed by radiologist and it has been estimated that only 15 to 30% of breast biopsy cases are proven to be cancerous [2].

Thus there is a significant necessity for a computer aided diagnosis system for providing radiologist with a low cost double reading (second opinion) without the drawback of fatigue or intra observer variability, improving the efficacy of screening program, and avoiding patient discomfort, cost, and probable breast scars which may cause diagnostic difficulty in future mammography examination, by avoiding unnecessary biopsies.

Computer aided detection has a record of investigation dating back to the 1960s where articles on computer analysis of radiographic images approved. It is important to distinguish computer aided detection versus computer aided diagnosis or classification. For computer aided detection a suspicious lesion is detected and localized (pinpointing) by some automated computer vision technique. Once the lesion has been detected by radiologist and /or some computer technique, computer aided diagnosis system then help the radiologist to classify that lesion or to make a patient management decision.

A. Cancer criteria

Benign masses defined by round, low density and smoothly sharply defined margins, while malignant masses have a high density, stellate, spiculated and poorly defined margins.

Benign microcalcifications: 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. The rule of malignancy that when the number of micro calcifications in a cluster is greater (usually more than 5); the likelihood of malignancy become greater. Typically, malignant micro calcifications present with a wide range in size, shape, and density.

B. Data source

The data collection that was used in our experiments was taken from the digital data base for screening mammography (South Florida University) it contain 2620, four view, mammography screening exams. The existing data in the collection include the location of the abnormalities in form of its boundary provided as chain code where the first two

values are the starting column and row of the lesion boundary while other numbers correspond to a specific direction on the X and Y coordinates.

II. BACKGROUND

The proposed system is built based on wavelet analysis of the region of interest and by applying the minimum distance and the voting K-Nearest Neighbor techniques for the classification stage. Here we introduced the theoretical background for both.

A. Wavelet Analysis

Wavelet analysis is the most recent solution to overcome the shortcoming of the Fourier transform. Wavelet is a waveform of limited duration and can be expressed as mathematical functions that cut up data into different frequency components (into shifted and scaled versions of the original or mother wavelet) and then study each component with a resolution matched to its scale.

The fundamental idea behind wavelet is to analyze according to scale. The spectrum is calculated each time it shifted and repeated many times with a slightly shorter or (longer) window every new cycle. So wavelet analysis allows the use of long time intervals where we want more precise low-frequency and shorter regions where we want high-frequency information [1].

Wavelet analysis is based on three properties: orthogonality, quadratic filter and filter bank. Two functions f and g are said to be orthogonal to each other if their inner product is zero.

$$\langle f(t), g(t)\rangle = \int_a^b f(t)g^*(t)dt = 0$$

The symbol * mean a convolution operation.

Dilation and translation of the mother function or analyzing function achieved as shown in equation:

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

The variables j and k are integers that scale and dilate the mother function φ to generate wavelets. The scale index j indicates the wavelet's width and the location index k gives its location or (translation).

In the two dimensions wavelet analysis, two dimensions scaling functions $\varphi(x, y)$ and three 2D wavelets are required. These wavelet functions measure intensity or gray level variations for image along different directions.

$\psi^H(x, y)$ responds to variation along columns (horizontal edge), $\psi^V(x, y)$ responds to variation along rows (vertical edges) and $\psi^D(x, y)$ measures variations along diagonals.

The discrete wavelet transform of image f(x, y) of size M x N is computed as follow:

We first define the scaled and translated basis functions:

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

$$\psi_{j,m,n}(x, y) = 2^{j/2}\psi(2^j x - m, 2^j y - n)$$

i: is a subscript that assumes values of H, V and D. then:

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

$$w_\psi^i(j_0, m, n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{x=0}^{N-1} f(x, y)\psi_\psi^i(x, y)$$

j_0: starting scale, we normally let it equal to zero.

$w_\varphi(j_0, m, n)$: define an approximation of f(x, y) at the scale and normally obtained by convolving signal with the low -pass filter.

$w_\psi^i(j, m, n)$: define horizontal, vertical and diagonal details normally obtained using high - pass filter.

B. Classification Stage

In order to assess the discriminative power of extracted features and classifiers,

Two statistical classification schemes were applied against the verified diagnosis for each case, minimum distance classification and K-nearest neighbor classification.

An important initial step of classification is to divide the data into two independent subsets, learn and test sets. This step is important to avoid the bias effects in the error estimation phase [3].

B1. The minimum distance classification:

This method assumes that the classes are similar in distribution and are linearly separable. Hence, the decision lines are allocated half way between the centers of clusters of different classes.

The algorithm works as follow:

Firstly, group the learn set into two supervised cluster according to their labels (malignant, and normal), representing the two pathologies of interest. Then, estimate the sample mean for each class by averaging the parameter set of the class. Then a test sample is classified by assigning it to the class which has the nearest mean vector.

Finally, error rate is estimated by the percentage of misclassified samples.

B2. The voting K-Nearest Neighbor (K-NN) classification:

K-nearest neighbor (K-NN) classifier distinguishes unknown patterns based on the similarity to known samples.

The K-NN algorithms computes the distances from an unknown patterns to every sample and select the K-nearest samples as the base for classification.

The unknown pattern is assigned to the class containing the most samples among the K-nearest samples.

III. METHEODOLOGY

The image histogram carries important information about the content of an image and can be used for discriminating the abnormal tissue from the local healthy background using texture features.

Visual characteristics currently used by radiologists to distinguish between malignant and benign masses include analysis of the contour of the mass, density, shape, size,…..etc, (Morphological features) [2].

The capabilities of Mat lab's Wavelet Toolbox were utilized in this study.We first import the mammograms of (LJPEG) format and the associated data files (as text files) which describe the abnormalities locations. As a preprocessing stage the mammograms gray levels is then mapped to its corresponding optical density for standardization purpose and contrast stretching performed. The software was prepared to localize the abnormalities using information associated within the data files. Then we able to present the standardize mammograms with the region of interest highlighted. The 64x64 pixel region of interest was then determined around the center of the abnormalities. Wavelet decomposition was applied over these regions and the statistical features and wavelet coefficients were then extracted from each wavelet detail at each level. These features were then presented to both the voting k-nearest and minimum distance classifiers to judge the normality and abnormality of the imaged tissue

The entire procedure of system development is presented in Fig (1). The following gives a detailed description of each step.

A. preprocessing stage:

Four different scanners types and models were used to create the mammography database, and since the gray levels in images acquired on different scanners will probably not correspond to the same optical density, a program to map these gray levels to its optical density values are used. All images were then full scale contrast stretched by linearly mapping the produced optical density values (in the range of −0.0903 to 4.4691) to grey levels in the range of (4095 to 0). Such a conversion provides a baseline correction for images digitized using different scanners. This allows us to run our image analysis software on data sets that were digitized on these four different scanners ensures the reliability of our extracted features and provides us with a reasonable standardization. Equations describing scanners optical density are provided within data set collection

B. Feature extraction:

Features have the capability of relating to image processing algorithms is preferred since it allows automatic extraction by computer and improved objectivity and reproducibility.

Using the chain codes, a lesion on a mammogram was identified and a rectangular bounding box of 64x64 pixels centered on the lesion called the region of interest (ROI) was selected

In this paper all features are extracted from regions of interest based on the wavelet decomposition. The features passed to the classification stage include the wavelet coefficients, and statistical features.

B1. Wavelet decomposition coefficients:

In this feature extraction stage, the wavelet decomposition applied on the region of interest using both the wavelet name as Daubechues (db4) and the function wmaxlev provided by mat lab toolbox to determine the maximum wavelet decomposition scale N, it helps to avoid unreasonable maximum scale values according to number of scales that contain irredundant information**.** The output of wavelet analysis are the decomposition vector C and corresponding book keeping matrix S, The vector C consist from horizontal, vertical, and diagonal detail coefficients and one approximation. The horizontal, vertical and diagonal detail was extracted from the wavelet decomposition structure [C, S]. These vectors were extracted at each scale from scale one to N+1. The coefficients vectors [H, V and D] for scale one to 4 were then normalized by dividing each vector by its maximum value. The result is that all vectors values become less than or equal one. Then we compute the energy for each vector by squaring every element in the vector.

Since high number of coefficients is produced we reduce these numbers by summing a predefined number of energy values together in a single number.

The produced values are then considered as features for the classification stage

B2. Statistical features:

Wavelet theory provides a powerful framework for multiresolution analysis, and it can be used for texture analysis. The wavelet transform is used to map the regions of interest into a series of coefficients, which constitute a multiscale representation of the ROIS [4].

In this paper and from each scale of the wavelet transform 11 statistical descriptors that include mean, standard deviation, and higher order statistics of intensity values are estimated for each ROIS.

Mean: $\mu = \sum_{k=1}^{N} f_k p_f(f_k)$,

Variance: $\sigma^2 = \sum_{k=1}^{N} (f_k - \mu)^2 p_f(f_k)$,

Skewness: $\mu_3 = \frac{1}{\sigma^3} \sum_{k=1}^{N} (f_k - \mu)^3 p_f(f_k)$,

Kurtosis: $\mu_4 = \frac{1}{4} \sum_{k=1}^{N} (f_k - \mu)^4 p_f(f_k)$,

Where N denotes the number of gray levels in the wavelet detail image at each scale. f_k is the kth gray level and $p_f(f_k) = n_k / n$, where n_k is the number of pixel with f_k gray level and n is the total number of pixels in the region.

The first percentile of the gray level distribution (P_1)

$$\sum_{j=0}^{p_1-1} h_j < \frac{1}{10} \leq \sum_{i=0}^{p_1} h_i$$

Where h_j is the value of the gray level histogram at gray level j. the first percentile represents the gray level at or below which lies 10% of the total number of pixels inside the ROI.

Some additional parameters derived using gray level co-occurrence matrix $p(i, j)$ to describe the gray level spatial

Inter-relationships. These parameters represent efficient measures of the gray level texture homogeneity.

Contrast= $\sum_{i,j}(i-j)^2.p(i,j)$,

Homogeneity= $p(i,j)/1+|i-j|$,

Entropy: $E=-\sum_{i,j}P(i,j).\log[\ p(i,j)]$,

Correlation= $\dfrac{\sum_{i,j}ijp(i,j)-m_i m_j}{s_i s_j}$,

Where $m_i=\sum_i i\sum_j p(i,j)$

$m_j=\sum_j j\sum_i p(i,j)$

$s_i^2=\sum_i i^2\sum_j p(i,j)-m_i^2$

$s_j^2=\sum_j j^2\sum_i p(i,j)-m_j^2$

Energy= $\sum_{i,j}p(i,j)^2$.

IV. RESULTS

Seventy two regions of interest extracted from 40 X-ray mammograms were used for evaluating the method.36 of these regions are known to be malignant while the remaining 36 regions are known to be normal.36 regions out of the total (72) were left for testing the system chosen to be 18 normal and 18 malignant. Different combinations of wavelet levels are investigated, level 1 to 4, level 1 to3 and level 2 to3. The results of both classifiers are listed in tables below.

In figure (1) the part titled as Center – boundary distances fluctuation is a result of the analysis of speculation. The graphs for cancer regions are found to exhibit more variation than benign one, measures of shape descriptors is also significant. A well designed system combining both classifiers and take into account the shape descriptor will take the advantage of each and expected to perform better.

Level 1 to 4

Table I
MINIMUM DISTANCE CLASSIFIRE

Malignant (training)	Normal (training)	Malignant (test)	Normal (test)
100%	77,78%	100%	55,56%

Level 2 to 3

Table II
MINIMUM DISTANCE CLASSIFIRE

Malignant (training)	Normal (training)	Malignant (test)	Normal (test)
100%	100%	100%	94,44%

Level 1 to 3

Table III
MINIMUM DISTANCE CLASSIFIRE

Malignant (training)	Normal (training)	Malignant (test)	Normal (test)
100%	100%	100%	94,44%

Level 1 to4

Table I
VOTING k-NN CLASSIFIER

k	Malignant (training)	Normal (training)	Malignant (test)	Normal (test)
1	100%	94,44%	100%	94,44%
3	100%	100%	100%	100%
5	100%	100%	100%	100%
7	100%	100%	100%	100%
9	100%	100%	100%	94,44%

Level 2 to 3

Table II
VOTING k-NN CLASSIFIER

k	Malignant (training)	Normal (training)	Malignant (test)	Normal (test)
1	100%	94,44%	100%	61,11%
3	100%	100%	100%	100%
5	100%	100%	100%	100%
7	100%	100%	100%	100%
9	100%	100%	100%	100%

Level 1 to 3

Table III
VOTING k-NN CLASSIFIER

k	Malignant (training)	Normal (training)	Malignant (test)	Normal (test)
1	100%	94,44%	100%	61,11%
3	100%	100%	100%	100%
5	100%	100%	100%	100%
7	100%	100%	100%	100%
9	100%	100%	100%	100%

REFRENCESES

1- Rafayah Mousa, Qutaishat Munib, Abdallah Moussa. "Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural". *Expert Systems with Applications 28 713–723. 2005.*

2- Maryellen L. Giger, Carl J. Vyborny, Robert A. Schmidt. "Computerized characterization of mammographic masses: analysis of speculation". The University of Chicago, MC2026, 5841 S. Maryland, Chicago, IL 60637, USA. *Cancer Letters 77 201-21 I. 1994*.

3- Yasser M. Kadah, Aly A. Farag, Jacek M. Zurada, Ahmed M.Badawi, Abou-Bakr M. Mousses. "Classification Algorithms for quantitative characterization of diffuse liver disease from ultrasound images".

4-H.D. Cheng, Xiaopeng Cai, Xiaowei Chen, Liming Hu, Xueling Lou"Computer-aided detection and classification of microcalcifications in mammograms: a survey". *Pattern Recognition 36 (2003) 2967 – 2991. 2003*

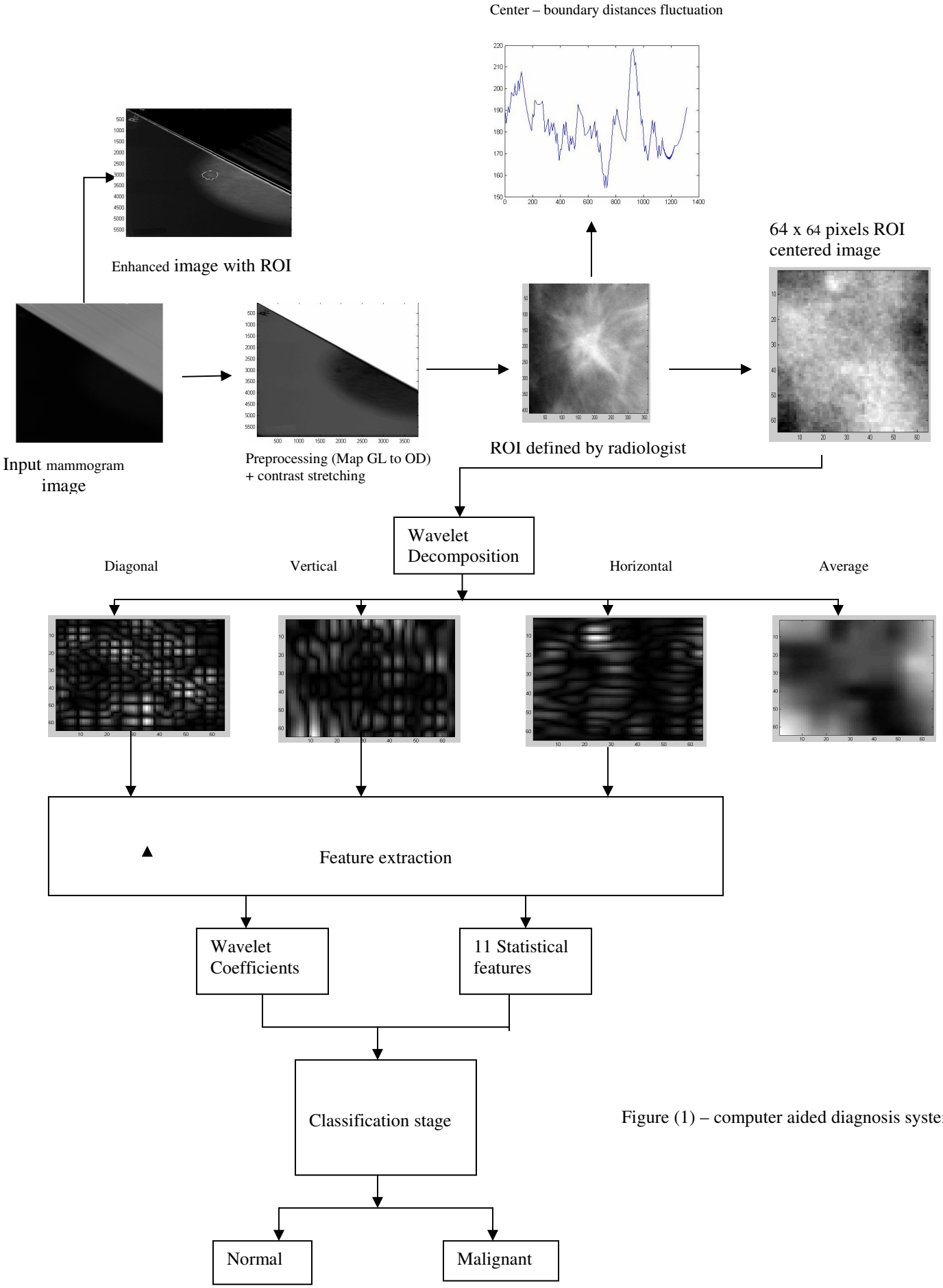

Figure (1) – computer aided diagnosis system