

Mass Detection in Mammograms

Ping-Sung Liao¹, Shu-Mei Guo², Nan-Sue Yu², Li-Chun Chen², San-Kan Lee^{3*}, and Chein-I Chang⁴

ABSTRACT

Many research efforts devoted to development of computer-aided-diagnosis (CAD) systems for mass detection have been focused on feature extraction/selection which is a crucial step in success of a CAD system. This paper investigates and evaluates many feature extraction techniques developed for mass detection in mammograms. In particular, two major techniques, texture spectrum and texture feature coding method are explored and a new feature extraction descriptor, called just noticeable difference (JND) is introduced. In order to improve accuracy for mass detection, the principal components analysis (PCA) and a new proposed genetic algorithm (GA) are used to select an optimal set of features that are fed to two neural network classifiers, backpropagation neural network (BPNN) and probabilistic neural network (PNN) for classification. The experimental results show that the proposed genetic algorithm outperforms the PCA in feature selection. The results also show that the best classification can be obtained by combining the proposed GA with a PNN classifier.

I. INTRODUCTION

Breast cancer is a second leading cause of cancerous diseases in women. However, the fatality can be greatly reduced through early diagnosis and detection. American Cancer Society strongly recommended that women over 40 years of age receive mammogram screening once every two years, and women over 50 years of age do every year. This is because it has been shown that mammography is the most effective screening modality for breast cancer detection. As a consequence, such a routine mammography screening may produce a huge amount of mammograms to be scrutinized by radiologists. Laurie Fajardo from University of Arizona's Tucson Breast Cancer Center indicated that every radiologist has to view 75 mammograms every day and only a few of them will be found unusual [1]. In addition to such heavy load of work for radiologists, signs of abnormality are usually too small or too subtle for them to pick up. For this reason, in order to improve the diagnostic accuracy and efficiency of

screening mammograms, computer-aided diagnosis (CAD) is introduced into the screening process to provide radiologists with a second opinion.

It is known that variability resulting from characteristics of breast abnormalities, such as diagnostic features, shapes, intensities and textures provide vital and crucial information in detection and classification of different types of breast tumors. Various approaches have also been developed along this line for feature extraction and selection. In 1989, Lai et al. used a template-matching method with circular templates to detect circumscribed mass, based on the fact that malignant tumors can be identified as approximately circular regions on mammograms [2]. Kobatake et al. developed a new filter called the iris filter to enhance only rounded opacities, and later in 1996, they came up with line skeleton and a modified Hough transform for detection of spicules which can be recognized as line patterns radiating from the center of tumor [3-6]. Mudigonda et al. used a polygonal model for a mass boundary, and generated features measuring the concavity or convexity of the boundary and the degree of spiculation for the classification of breast masses [7-8]. While continuous research efforts on detection of breast tumors have resulted in many more new feature descriptors, how to select effective feature descriptors among such an overwhelming set of features descriptors becomes increasingly vital and crucial since not all extracted features provide desired and useful information. Therefore, an effective feature selection method must be taken into consideration in a designed CAD system. Sahiner et al. showed that a genetic algorithm for feature selection could work very effectively in selecting a useful subset from over 500 features, which are mainly texture-related features [10].

In this paper, a variety of features extraction techniques are chosen to make our feature space thorough, so that we can evaluate the performance of principal components analysis (PCA) and a genetic algorithm combined with two neural network classifiers in recognizing all kinds of breast abnormalities from three types of breast tissue samples. In the following section, Section II, we will first give a description of the methods proposed in this study. They are organized into four parts, segmentation, feature extraction, feature selection and classification. Then, the experimental results and conclusions will be presented in Section 3 and 4, respectively.

II. SYSTEM ARCHITECTURE

Paper submitted 08/15/11; revised 11/15/11; accepted 12/30/11.

*Corresponding Author: San-Kan Lee.(E-mail: sklee@vghcc.gov.tw)

¹ Department of Electrical Engineering, Cheng Shiu University, Kaohsiung, Taiwan

² Department of Computer Science and Information Engineering National Cheng Kung University, Tainan, Taiwan

³ Taichung Veterans General Hospital, Taichung, Taiwan

⁴ Remote Sensing Signal and Image Processing Laborator, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250, U.S.A

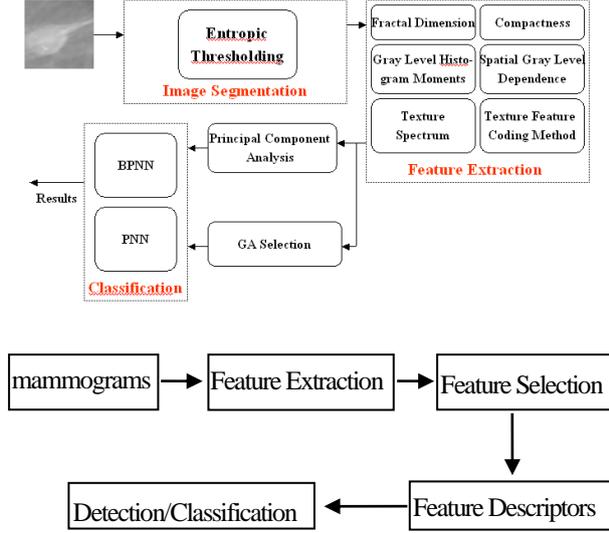


Figure 1. System Architecture

1. Image Segmentation — entropic thresholding and morphological processing

First, we calculate the co-occurrence matrix $W_{L \times L}$ from an $L \times L$ image denoted by $I(i,j)$ where the element of W at row i and column j , W_{ij} denotes the number of pixels with gray level i while pixels in their horizontally right and/or vertically below positions having gray level j . For each pair of gray level (i, j) , define the joint probability $p(i, j)$ as follows:

$$p(i, j) = \frac{W_{ij}}{\sum_{i=1}^L \sum_{j=1}^L W_{ij}} \quad (1)$$

$$W_{ij} = \sum_{l=1}^M \sum_{k=1}^N \delta(l, k) \quad (2)$$

with $\delta(l, k)$ defined as follows:

$$\delta(l, k) = \begin{cases} 1 & \text{if } I(l, k) = i, I(l, k+1) = j \text{ and/or } I(l, k) = i, I(l+1, k) = j \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Using a threshold denoted by T , we can segment the image $I(i,j)$ into two parts; one is called the Object with gray levels larger than the threshold T , and the other is the Background with gray levels smaller than T . Next, we further partition the co-occurrence matrix W by the threshold T into four regions, which are Background to Background (BB), Background to Object (BO), Object to Background (OB) and Object to Object (OO). Finally, we calculate probabilities for each region from which we can obtain the entropy of each region as follows:

$$H_{BB}(T) = - \sum_{i=1}^T \sum_{j=1}^T p_{BB}(i, j) \log p_{BB}(i, j) \quad (4)$$

$$H_{BO}(T) = - \sum_{i=1}^T \sum_{j=T+1}^L p_{BO}(i, j) \log p_{BO}(i, j) \quad (5)$$

$$H_{OB}(T) = - \sum_{i=T+1}^L \sum_{j=1}^T p_{OB}(i, j) \log p_{OB}(i, j) \quad (6)$$

$$H_{OO}(T) = - \sum_{i=T+1}^L \sum_{j=T+1}^L p_{OO}(i, j) \log p_{OO}(i, j) \quad (7)$$

By virtue of (4-7) and [11] the local entropy (LE), $H_{LE}(T)$ can be defined by

$$H_{LE}(T) = H_{BB}(T) + H_{OO}(T) \quad (8)$$

A method that finds a gray level value, T_{LE} maximizing (8), expressed by

$$T_{LE} = \arg\{\max_T H_{LE}(T)\} \quad (9)$$

is referred to as local entropy (LE) method

Since it is often the case that the image processed by entropy thresholding may contain some noises such as light dots in a dark background or dark elements on a large white area, morphological opening and closing operations are generally applied alternately to eliminate those noisy effects [12] so that the segmented regions can be processed in the next stage.

2. Feature Extraction

Feature extraction is one of key factors to success in mass detection and classification. Many techniques have been developed for this purpose. In this section we investigate six categories of feature extraction techniques, fractal dimension [13], compactness [9-14], gray level histogram [15], spatial gray level co-occurrence dependence, texture spectrum [16] and texture feature coding method (TFCM) [17]. Since the first two categories, fractal dimension and compactness, involves only one parameter, there is no need to deliberate here. Only the remaining four categories are discussed as follows.

2.1 Gray Level Histogram

Gray level histogram-based methods include mean, variance skewness and kurtosis.

2.2 Spatial gray level dependence category

11 feature parameters, energy, inertia, intensity, entropy, contrast, correlation, inverse difference moment, sum of squares variance, sum average, sum entropy, and difference entropy are calculated from co-occurrence matrix [9, 13].

By adjusting the distance d between two pixels, and the angle θ one pixel deviated from another in calculating co-occurrence matrix, different matrices are obtained. In this study, we have 16 different co-occurrence matrices according to $d = 1, 2, 3, \text{ and } 4$ combined with $\theta = 0^\circ, 45^\circ, 90^\circ, \text{ and } 135^\circ$ respectively. Therefore, we have 176 features in total in this category.

E_1 (P_1, V_1)	E_2 (P_2, V_2)	E_3 (P_3, V_3)
E_8 (P_8, V_8)	(P_0, V_0)	E_4 (P_4, V_4)
E_7 (P_7, V_7)	E_6 (P_6, V_6)	E_5 (P_5, V_5)

Figure 2. Relation of pixels in 8-neighbor connectivity

2.3 Texture Spectrum

Texture spectrum category, given a pixel and its neighborhood under a 3x3 mask, a vector $V = \{V_0, V_1, \dots, V_8\}$ represents the gray levels of this 3x3 neighborhood, where V_0 is the gray level of the center pixel, and V_1, V_2, \dots, V_8 are the gray levels of its 8-neighbors starting in the clockwise direction from the upper left position, as in Fig. 2.

For the given 3x3 neighborhood, the corresponding Texture Unit (TU) is defined by $TU = \{E_1, E_2, \dots, E_8\}$,

where E_i represents the gray-level relationship between a pixel and its i-th neighbor, which is defined separately as Eq. (10) in our first coding method [18] and as Eq. (11) in the second coding method [15]. The reason why we adopt two coding methods simultaneously is that the first coding method is good for abnormality detection while the second coding method is good for classification from our experimental analysis.

$$E_i = \begin{cases} 0, & \text{if } V_i < V_0 \\ 1, & \text{if } V_i = V_0 \\ 2, & \text{if } V_i > V_0 \end{cases} \quad \text{for } i=1,2,\dots,8 \quad (10)$$

$$E_i = \begin{cases} 0, & \text{if } (V_0 - V_i) > s \times JND(V_0) \\ 1, & \text{if } |V_i - V_0| \leq s \times JND(V_0) \\ 2, & \text{if } (V_i - V_0) > s \times JND(V_0) \end{cases} \quad \text{for } i=1,2,\dots,8 \quad (11)$$

Where

$$JND(V_0) = \begin{cases} 17 \times [1 - (\frac{\bar{g}_8}{127})^{1/2}] + 3, & \text{if } \bar{g}_8 \leq 127 \\ (3/128) \times (\bar{g}_8 - 127) + 3, & \text{if } \bar{g}_8 > 127 \end{cases} \quad (12)$$

with \bar{g}_8 being the average gray level of the eight-neighbors and s being the weight which varies with different images. According to our experiments, 0.3-0.5 are found to be most suitable for s in mammographic detection.

From the above definition, there are three possible values for each element of TU . Therefore, there are $3^8 = 6561$ combinations of texture units. We then define the texture unit number N_{TU} by the binary coding theory as below :

$$N_{TU} = \sum_{i=1}^8 E_i \times 3^{i-1} \quad (13)$$

8 feature parameters are obtained through NTU, which are Black-Whit Symmetry, Geometric Symmetry, Degree of Direction, Micro Horizontal Structure, Micro Vertical Structure, Micro Diagonal Structure 1, Micro Diagonal Structure 2, and Central Symmetry [13]. Thus, we have 16 features extracted in total by the two coding methods.

2.4 Texture Feature Coding Method (TFCM)

The TFCM was originally proposed in [17] and can be viewed as an extension of the texture spectrum. It considers three consecutive pixels with certain specific directions as shown by dotted lines in Fig. 3(a-b), and calculates gradient changes in gray levels of the two successive adjacent pixels among these three pixels. In other words, for a given seed pixel labeled by X_0 centered in a 3x3 mask, two types of neighborhood connectivities are of interest. One is called the first-order 4-neighbor connectivity consisting of the pixels labeled by X_1, X_3, X_5 and X_7 in the horizontal and vertical direction of the seed pixel X_0 shown in Fig. 3(a). The other is called the second order 4-neighbor connectivity consisting of the pixels labeled by X_2, X_4, X_6 and X_8 in the diagonal and anti-diagonal lines of the seed pixel X_0 shown in Fig. 3(b).

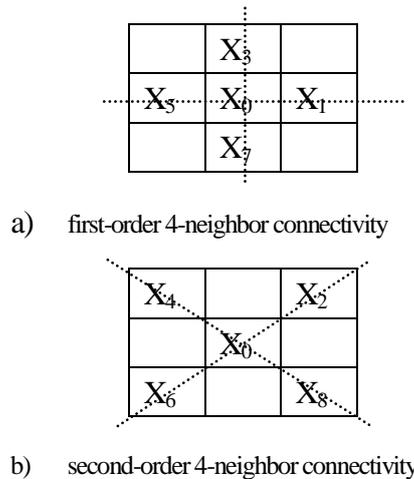


Figure 3. Two types of 4-neighbor connectivities

Furthermore, let C be represent the gray level of the seed pixel, X_0 in the 3×3 mask, and $\{a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4\}$ be the gray levels of the eight neighboring pixels as shown in Fig. 4.

a_4	a_2	a_3
a_1	C	b_1
b_3	b_2	b_4

Figure 5. a 3x3 mask

Then the gray level variations among three consecutive

pixels along specific directions indicated by dotted lines in Fig. 3(a-b) can be divided into 4 types as follows :

$$(I) (|a_i - C| \leq \Delta) \cap (|C - b_i| \leq \Delta)$$



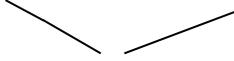
(II)

$$[(|a_i - C| \leq \Delta) \cap (|C - b_i| \geq \Delta)] \cup [(|a_i - C| \geq \Delta) \cap (|C - b_i| \leq \Delta)]$$



(III)

$$[(a_i - C > \Delta) \cap (C - b_i > \Delta)] \cup [(C - a_i > \Delta) \cap (b_i - C > \Delta)]$$



(IV)

$$[(a_i - C > \Delta) \cap (b_i - C > \Delta)] \cup [(C - a_i > \Delta) \cap (C - b_i > \Delta)]$$



where the parameter Δ is included to specify a desired gray level tolerance.

In light of Δ , two types of coding methods are proposed. One is to use a fixed gray level tolerance with a range from 1 to 3 [14,16]. According to our extensive experiments, the gray level 3 seems to yield the best performance. The other is to make Δ adaptive that can be adjusted in accordance with the concept, called Just Noticeably Different (JND), introduced in [19]. With the use of JND, an adaptive is defined by

$$\Delta = s \times JND(V_0) \quad (14)$$

It has been shown that the TFCM is sensitive to the selection of the which must be adjusted locally and adaptively and an appropriately selected is the key to success of the TFCM.

Given a pixel at (x, y) , we can further introduce a pair of parameters, (α, β) where α represents the combination of the gray-level variation along the first scan line in first order connectivity with the gray-level variation along the second scan line of second order connectivity, and β represents the combination of the gray-level variation along the second scan line in first order connectivity with the gray-level variation along the first scan line in second order connectivity. In terms of the type of the gray-level variation, the possible values of α or β are coded as in Table 2.

Table 2. The combined variations of first order connectivity with second order connectivity

Type	(I)	(II)	(III)	(IV)
second order	(I)	(II)	(III)	(IV)
	1	2	3	4
	(II)	5	6	7
	(III)	8	9	
	(IV)	10		

According to Table 2 a Texture Feature Number (TFN) can be defined for a given pixel (x, y) as follows.

$$TFN(x, y) = \alpha(x, y) \times \beta(x, y) \quad (15)$$

Although the value of the TFN lies within the range from 0 to 100 by its definition, there are only 42 possible values can be actually produced by (15). In this case, we re-assign numbers 0~41 to the 42 TFNs. By means of these 42 TFNs we can calculate 7 feature parameters, Coarseness, Homogeneity, Mean Convergence, Variance, Entropy, Similarity, and Regularity [18]. Finally, we have 14 features extracted for this category.

Finally, Table 1 summarizes and lists the 212 features from six categories, fractal dimension, compactness, gray level histogram, spatial gray level dependence, texture spectrum and texture feature coding method (TFCM), and the equations that are used to generate these features along with their corresponding references.

Table 1. Features

Categories	Features	Equations
Fractal Dimension	Fractal Dimension	[13]
Compactness	Compactness	$1 - P^2 / A$
GrayLevel Histogram	Mean	$\mu = \sum_{k=0}^{255} k \cdot p(k)$
	Variance	$\mu_2 = \sum_{k=0}^{255} (k - \mu)^2 \cdot p(k)$
	Skewness	$\mu_3 = \frac{1}{\sigma^3} \sum_{k=0}^{255} (k - \mu)^3 \cdot p(k)$
	Kurtosis	$\mu_4 = \frac{1}{4} \sum_{k=0}^{255} (k - \mu)^4 \cdot p(k) - 3$
Spatial Gray Level Dependence	Energy	$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [p_{d,\theta}(i, j)]^2$
	Inertia	$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [p_{d,\theta}(i, j) \times (i - j)^2]$
	Intensity	$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [i \times j \times p_{d,\theta}(i, j)]$
	Entropy	$-\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{d,\theta}(i, j) \times \log p_{d,\theta}(i, j)$
	Contrast	$\sum_{n=1 \& \forall i, j}^{L-1} n^2 \times p_{d,\theta}(i, j)$ $ i - j = n$
	Correlation	$\frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} i \times j \times p_{d,\theta}(i, j) - \mu_x \times \mu_y}{\sigma_x \times \sigma_y}$
	Inver Difference Moment	$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{1}{1 + (i - j)^2} \times p_{d,\theta}(i, j)$
Sum of Squares Variance	$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu_x)^2 \times p_{d,\theta}(i, j)$	

	Sum Average	$\sum_{i=0}^{2L-2} i \times p_{x+y}(i)$
	Sum Entropy	$-\sum_{i=0}^{2L-2} p_{x+y}(i) \times \log(p_{x+y}(i))$
	Difference entropy	$-\sum_{i=0}^{L-1} p_{x-y}(i) \times \log(p_{x-y}(i))$
Texture Spectrum	Black-White Symmetry	$BWS = \frac{\sum_{i=0}^{3279} S_1(i) - S_1(3281+i) }{\sum_{i=0}^{6560} S_1(i)}$
	Geometric Symmetry	$GS = \frac{\sum_{i=0}^{6560} S_j(i) - S_{j+4}(i) }{4 \sum_{j=1}^4 \sum_{i=0}^{6560} S_j(i)}$
	Degree of Direction	$DD = \frac{\sum_{m=1}^3 \sum_{n=m+1}^4 S_m(i) - S_n(i) }{6 \sum_{m=1}^3 \sum_{n=m+1}^4 \sum_{i=0}^{6560} S_m(i)}$
	Micro Horizontal Structure	$MHS = \sum_{i=0}^{6560} S_1(i) \times HM(i)$ where $HM(i) = P(1,2,3) \times P(5,6,7)$
	Micro Vertical Structure	$MVS = \sum_{i=0}^{6560} S_1(i) \times VM(i)$ where $VM(i) = P(1,7,8) \times P(3,4,5)$
	Micro Diagonal Structure 1	$MDS1 = \sum_{i=0}^{6560} S_1(i) \times DM1(i)$ where $DM1(i) = P(1,2,8) \times P(4,5,6)$
	Micro Diagonal Structure 2	$MDS2 = \sum_{i=0}^{6560} S_1(i) \times DM2(i)$ where $DM2(i) = P(2,3,4) \times P(6,7,8)$
	Central Symmetry	$CS = \sum_{i=0}^{6560} S_1(i) \times [K(i)]^2$
Texture Feature Coding Method	Coarseness	$P_{\Delta}(n) = \frac{N_{\Delta}(n)}{N}$ $n \in \{0, 1, 2, \dots, 41\}$
	Homogeneity	$\sum_{\Delta} P_{\Delta}(0)$
	Mean Convergence	$MC = \sum_{n=0}^{41} \frac{ n \cdot P_{\Delta}(n) - \mu_{\Delta} }{\sigma_{\Delta}^*}$
	Variance	$\sum_{n=0}^{41} (n - \mu_{\Delta}^*)^2 \cdot P_{\Delta}^*(n)$
	Entropy	$P_{TFN_{\Delta, d, \theta}}(i, j) = \frac{W_{TFN_{\Delta, d, \theta}}(i, j)}{\sum_{i=0}^{41} \sum_{j=0}^{41} W_{TFN_{\Delta, d, \theta}}(i, j)}$
	Similarity	$\sum_{i=0}^{41} (P_{TFN_{\Delta, d, \theta}}(i, i))^2$
	Regularity	$\sum_{i=0}^{41} \sum_{j=0}^{41} \frac{P_{TFN_{\Delta, d, \theta}}(i, j)}{1 + (i - j)^2}$

3. Feature Selection

As noted in the previous section, a significant number of features can be generated from different feature extraction techniques and also be used for detection and classification. Apparently, not all features are useful and effective. Therefore, a follow-up task is to select an optimal set of features that meet our need. Principal Components Analysis (PCA) [20] has been widely used in dimensionality reduction. It is done by taking the largest q eigenvalues of the covariance matrix formed by the original feature vectors. Each principal component y_i is obtained by linearly combining the original features, where x is the original feature vector and v_i is the eigenvector of the covariance matrix.

$$y_1 = \mathbf{v}_1^T \mathbf{x} = v_{11}x_1 + v_{12}x_2 + \dots + v_{1N}x_N$$

$$y_2 = \mathbf{v}_2^T \mathbf{x} = v_{21}x_1 + v_{22}x_2 + \dots + v_{2N}x_N$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$y_N = \mathbf{v}_N^T \mathbf{x} = v_{N1}x_1 + v_{N2}x_2 + \dots + v_{NN}x_N$$

However, PCA transforms the original variables into new components which are linear combinations of the original variables. Sometimes it is better to simply choose a subset of the original features in certain applications rather than to linear transform them. Therefore, as an alternative, a genetic algorithm (GA) [21] is developed for feature selection.

The GA consists of five procedures, which are encoding, initial Population, fitness function, genetic operator, and stop criterion.

(a) Encoding :

The number of bits in a chromosome is equal to the total number of the acquired features, and each bit corresponds to an individual feature extracted from the region of suspicion. A specific feature is selected within a chromosome if the corresponding bit is set to 1, otherwise, it's ignored.

(b) Initial Population :

The number of the population is equal to the total number of chromosomes. Even though GA converges more quickly as the population becomes larger, it cost more computational time as well. So, in our experiment, the population at each generation is kept constant at 30. In each chromosome, 10 random bits are initialized to 1 and the rest of them are 0.

(c) Fitness Function :

The fitness function is the core of a GA. Each chromosome is evaluated by the fitness function, and it is determined by the value that the function generates which chromosome is to be chosen to crossover. Here, a new fitness function is proposed as Eq. (16) :

$$f(c) = \left(\frac{\sum_{i=1}^{total} b_i (\mu_{i,mass} - \mu_{i,normal})^2}{\sum_{i=1}^{total} b_i \sigma_{i,mass}^2 + \sum_{i=1}^{total} b_i \sigma_{i,normal}^2} \right) - p(x) \quad (16)$$

$$where \ p(x) = \begin{cases} 0 & \text{if } 7 \leq \sum_{i=1}^{total} b_i \leq 13 \\ \infty & \text{otherwise} \end{cases}$$

where c is a chromosome, and b_i is the i -th bit of chromosome. $\mu_{i,mass}$ and $\sigma_{i,mass}^2$ are the mean and variance of the i -th feature extracted from all ROIs of masses. The same as $\mu_{i,normal}$ and $\sigma_{i,normal}^2$, the mean and variance of the i -th feature extracted from all ROIs of normal tissues. The fitness function increases as $(\mu_{i,mass} - \mu_{i,normal})^2$ increases. Feature parameter with large difference between $\mu_{i,mass}$ and $\mu_{i,normal}$ is advantageous for differentiating masses from normal tissues. On the other hand, when either $\sigma_{i,mass}^2$ or $\sigma_{i,normal}^2$ is large, it means that the specific feature exhibits too much variation in describing certain characteristics either of abnormality or of normality to be a good choice in our feature set. In this way, larger $f(c)$ indicates that the corresponding features within the chromosome are more effective in detecting masses. Moreover, the numerator is normalized by $\sum_{i=1}^{total} b_i$ and the $p(x)$ is manipulated so that we can further confine the number of features being selected to the range from 7 to 13, which is a handy finding from the experiment with PCA.

(d) Genetic Operators :

Genetic operators fall into three categories, (1)Parent Selection (2)Crossover (3)Mutation. In parent selection, Roulette wheel selection is used for reproduction, where each parent chromosome in a generation is selected with a probability proportional to its fitness evaluation. After two parents p_1 and p_2 are selected, we use one-point crossover, where a random crossover point in a chromosome is chosen, and each of the parent chromosomes were split into left and right strings at this crossover point. Offspring is generated by combining the left string of p_1 with the right string of p_2 , and vice versa. Mutation was applied randomly to a certain bit of the chromosome in the new generation by setting this bit to 1 if it's 0 and vice versa.

(e) Stop Criterion :

In our experiments, once the fitness reaches our expected value, the process will be terminated.

4. Classifiers

Neural networks have been widely used for its efficiency in solving non-linear problems. We use multilayer neural networks as our classifiers. The multilayer networks are composed by the input layer, the output layer and the hidden layer connecting the first two. The input layer consists of input nodes and, here each feature is represented by one node. In our experiment, we use one node at the output layer, which will discriminate between malignant tumors and benign tissues. In the hidden layer, there could be more than one layer according to different network topologies, and the connections are characterized by weights, which will be optimized through proper training. The neural network learning can be either supervised or unsupervised.

In this paper, we adopt two supervised network architectures, Probabilistic Neural Network (PNN) and Back-Propagation Neural Network (BPNN). The PNN is a statistical Bayesian classifier using Gaussian distributions and it is a network with no need of training. The BPNN is a three-layer feed-forward neural network using sigmoid activation function and the steepest descent method for error correction. In the literature, the BPNN is probably the most widely used neural network.[22-23].

III. EXPERIMENTAL RESULTS

The database used for our experiments is the MIAS Minimammographic Database [24] provided by the Mammographic Image Analysis Society (MIAS). There are 207 normal mammograms compared to 115 mammograms which contain abnormal tissues. Three classes are considered in accordance with breast parenchyma, dense-glandular, fatty and fatty-glandular. Among 207 normal mammograms are 76 dense-glandular, 66 fatty and 65 fatty-glandular. As for 115 abnormal mammograms there are 36 dense-glandular, 40 fatty and 39 fatty-glandular with abnormality classified into six categories: well-defined/circumscribed masses, spiculated masses, architectural distortion, asymmetry, calcification and other/ill-defined masses. Since we were only interested in mass detection, those mammograms with calcifications were eliminated from our experiments. Moreover, in terms of breast parenchyma, we partitioned breast tissue samples into three types, dense-glandular, fatty and fatty-glandular, and the experiments were conducted to classify breast masses according to these three types. Abnormal samples were extracted from the square area in which they are located in the center and the radius of masses circled by the MIAS. Additionally, we also sampled two types of normal tissues, one of which was extracted from area near abnormality, named type A. Since they are close to abnormality, the normal samples will look similar to abnormal tissues and they are not easy to detect. The reason

for using type A is to test whether the proposed system can perform well under worse scenarios. The other type, named type B was extracted from normal mammograms, which are clear from any masses.

After extracting all the desired features, we normalize them into the interval [0, 1], and run the GA for 100 times to accumulate the number of times a feature has been selected. The more frequently the feature is selected, the more significant the feature is. From our experimental results, the highest frequency of a feature selected was approximately 7 ~ 9 times, as tabulated in Table 3.

Table 3 Optimal Feature Selection by GA

	Optimal Feature Set
Dense-glandular and type B	Skewness, Micro Horizontal Structure, Micro Vertical Structure, Micro Diagonal Structure 1, Micro Diagonal Structure 2, Central Symmetry, JND Micro Vertical Structure, JND Micro Diagonal Structure 1
Dense-glandular and type A	Micro Horizontal Structure, Micro Vertical Structure, Micro Diagonal Structure 1, sum average(d=4, 0 ⁰ , 45 ⁰ , 135 ⁰), JND Black-Whit Symmetry
Fatty and type B	Micro Vertical Structure, Central Symmetry, JND Micro Horizontal Structure, JND Micro Vertical Structure, JND Micro Diagonal Structure 1, JND Micro Diagonal Structure 2, JND Central Symmetry
Fatty and type A	Micro Vertical Structure, Central Symmetry, JND Micro Horizontal Structure, JND Micro Vertical Structure, JND Micro Diagonal Structure 1, JND Micro Diagonal Structure 2, JND Central Symmetry
Fatty-glandular and type B	Intensity(d=3,4, 0 ⁰ , 45 ⁰), Intensity(d=4, 90 ⁰ , 135 ⁰), JND Micro Horizontal Structure, JND Micro Vertical Structure, JND Micro Diagonal Structure 1
Fatty-glandular and type A	Micro Horizontal Structure, Micro Vertical Structure, Micro Diagonal Structure 1, Central Symmetry, JND Micro Horizontal Structure, JND Micro Vertical Structure, JND Micro Diagonal Structure 1

Finally, at the classification stage, we randomly choose half of the abnormal samples, with the same quantity's normal samples for training our classifiers. The remaining samples are then used for testing. Since the training samples and testing samples selection are randomized, the classification procedure is repeated for 100 times with different samples every time. The results are then recorded in terms of True Positive Number (TPN), True Negative Number (TNN), False Negative Number (FNN), and False

Positive Number (FPN), and three ratios are derived from the four parameters to approximate correction rate :

- (1) Detection Rate : $DR = TPN / N_p$
- (2) False Alarm Rate : $FAR = FPN / N_n$
- (3) Correct Classification Rate : $CR = (TPN + TNN) / (N_p + N_n)$

where $N_p = TPN + FNN$ (N_p is the amount of the testing abnormal samples), and $N_n = TNN + FPN$ (N_n is the amount of testing normal samples).

From Tables 4-5, it shows that the performance with the fatty samples is the best among the three types of breast tissues, and no matter whether the normal samples are type A or type B, the CR reaches around 90%. When comparing the efficacy of GA with that of PCA in mass detection, for whatever the type of breast tissues is, GA is better than PCA. When we look further into the results from the type A of normal samples and from the type B of normal samples, the CR with B type is better than that with type A. It goes without saying that using normal samples extracted near masses impairs the recognition accuracy due to the feature similarity with abnormal tissues. Notwithstanding, the performance with type A is still great with over 80% CR in our system, and the average false alarm rate is around 8%.

Table 4 The Result of Masses and B Type

(a) Result of Dense Glandular							
	TPN	FNN	TNN	FPN	DR	FAR	CR
BPNN (PCA)	8.86	4.14	47.21	15.79	68.15%	25.06%	73.78%
BPNN (GA)	8.79	4.21	59.72	3.28	67.62%	5.21%	90.14%
PNN (PCA)	10.22	2.78	49.74	13.26	78.62%	21.05%	78.89%
PNN (GA)	10.04	2.96	56.38	6.62	77.23%	10.51%	87.39%
(b) Result of Fatty							
	TPN	FNN	TNN	FPN	DR	FAR	CR
BPNN (PCA)	14.82	3.18	42.97	5.03	82.33%	10.48%	87.56%
BPNN (GA)	16.45	1.55	47.96	0.04	91.39%	0.08%	97.59%
PNN (PCA)	15.53	2.47	45.38	2.62	86.28%	5.46%	92.29%
PNN	16.84	1.16	48	0	93.56%	0%	98.24%

(GA)							
(c) Result of Fatty Glandular							
	TPN	FNN	TNN	FPN	DR	FAR	CR
BPNN (PCA)	9.77	5.23	39.41	10.59	65.13%	21.18%	75.66%
BPNN (GA)	12.22	2.78	41.16	8.84	81.47%	17.68%	82.12%
PNN (PCA)	12.96	2.04	41.99	8.01	86.4%	16.02%	84.54%
PNN (GA)	12.32	2.68	45.52	4.48	82.13%	8.96%	88.98%

Table 5 The Result of Masses and Type A

(a) Result of Dense Glandular							
	TPN	FNN	TNN	FPN	DR	FAR	CR
BPNN (PCA)	8.67	4.33	9.79	3.21	66.69%	24.69%	71%
BPNN (GA)	10.81	2.19	10.37	2.63	83.15%	20.23%	81.46%
PNN (PCA)	11.41	1.59	10.3	2.7	87.77%	20.77%	83.5%
PNN (GA)	10.49	2.51	10.75	2.25	80.69%	17.31%	81.69%
(b) Result of Fatty							
	TPN	FNN	TNN	FPN	DR	FAR	CR
BPNN (PCA)	15.27	2.73	15.75	2.25	84.83%	12.5%	86.17%
BPNN (GA)	16.44	1.56	17.96	0.04	91.33%	0.22%	95.56%
PNN (PCA)	15.65	2.35	16.31	1.69	84.94%	9.39%	88.78%
PNN (GA)	16.83	1.17	18	0	93.5%	0%	96.75%
(c) Result of Fatty Glandular							
	TPN	FNN	TNN	FPN	DR	FAR	CR
BPNN (PCA)	8.49	6.51	11.12	3.88	56.6%	25.87%	65.37%

BPNN (GA)	10.46	4.54	14.29	0.71	69.73%	4.73%	82.5%
PNN (PCA)	11.69	3.31	9.32	5.68	64.73%	37.87%	63.43%
PNN (GA)	11.46	3.54	14.12	0.88	76.4%	5.87%	85.27%

IV. CONCLUSIONS

Our goal is to find out an optimal feature selection technique for detecting all five different masses at the same time. To accomplish that by depending on features of certain type can not be very effective. Therefore, in this paper, we extract 212 features, where a new category of feature descriptor, the JND-related features, is proposed. It turns out to be a very useful feature descriptor, which is proved by the high frequency they are selected by the proposed GA. On table 3 and table 4, it indicates that GA combined with PNN comes up with the best result. In conclusion, disregarding the type of breast tissues, GA outperforms PCA.

It is worth mentioning that the proposed system also achieves good performance while working on A type of normal samples, which further proves that the proposed system have great capability in mass detection despite homogeneity between normal and abnormal samples.

REFERENCES

- [1] K. F. Schmidt, "A better breast test," Science News, 143, pp. 392-393, 1993.
- [2] S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," IEEE Trans. Med. Imag., Vol. 8, pp.377-386, Oct. 1989.
- [3] K. Matsumoto, H. R. Jin and H. Kobatake, "Detection method of malignant tumors in DR images-Iris filter," Trans. Inst. Electron. Inform. Communi. Eng. D-II, Vol. J75-D- II, No. 3, pp. 663-670, 1992.
- [4] H. Kobatatke and M. Murakami, "Adaptive filter to detect rounded convex regions: Iris filter," Proc. Int. Conf. Pattern Rec., Vol. 2, pp. 340-344, 1996.
- [5] H. Kobatatke, M. Murakami, S. Nawano, N. Nakajima, and H. Takeo, "Accurate detection of malignant tumors on mammogram," Proc. CAR'96, pp. 1076-1081, 1996.
- [6] H. Kobatatke and S. Hashimoto, "Convergence index filter for vector field." IEEE Trans. Image Processing, Vol. 8, pp. 1029-1038, 1999.
- [7] N. R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels, "Concavity and convexity analysis of mammographic masses via an iterative

- segmentation algorithm,” Proc. IEEE Canadian Conference on Electrical and Computer Engineering, Edmonton, AB, pp 1489-1494, May 1999.
- [8] R. M. Rangayyan, N. R. Mudigonda, and J. E. L. Desautels, “Boundary modeling and shape analysis methods for classification of mammographic masses,” *Med. Biol. Eng. Comput.*, Vol. 38, No. 5, pp.487-496, 2000.
- [9] I.Christoyianni, E. Dermatas, G. Kokkinakis, “Neural classification of abnormal tissue in digital mammography using statistical features of the texture,” *Proc. ICECS '99*, Vol 1, pp.117-120, 1999.
- [10] Berkman Sahiner, Heang-Ping Chan, Datong Wei, Nicholas Pertick, Mark A. Helvie, Dorit D. Adler, and Mitchell M. Goodsitt, “Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue,” *Medical Physics*, Vol. 23, No. 10, pp. 1671-1684, October 1996.
- [11] N.K. Pal and S.K. Pal, “Entropic thresholding,” *Signal Processing*, Vol. 16, pp.97-108, 1989.
- [12] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, Prentice Hall, New Jersey, 2002.
- [13] A.P Pentland, “Fractal based description of natural scenes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 6, pp.661-674, 1984.
- [14] H. Kobatake, M. Murakami, H. Takeo and S. Nawano, “Computerized detection of malignant tumors on digital mammograms,” *IEEE Trans. Med. Imag.*, Vol. 18, No. 5, pp. 369-378, May 1999.
- [15] B. Aldrich and M. Desai, “Application of spatial grey level dependence methods to digitized mammograms,” *Proc of IEEE Southwest Sym. On Image Analysis and Interpretation*, pp. 100-105, 1994.
- [16] D.C. He and L. Wang, “Texture features based on texture spectrum,” *Pattern Recognition*, vol. 24, no. 5, pp. 391-399, 1991.
- [17] M.-H. Horng, “Texture feature coding method for texture classification,” *Optical Engineering*, vol. 42, no. 1, pp. 228-238, Jan. 2003.
- [18] Shu-Mei Guo, Ping-Sung Liao, Yu-Chain Liao, Sheng-Chih Yang, Pau-Choo Chung, San-Kan Lee and Chein-I Chang, “Mass Detection in Mammography Using Texture Analysis,” *Chinese Journal Radiology*, Vol. 28, No. 3, pp. 149-157, June 2003.
- [19] A. K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [20] L. J. Cao and W. K. Chong, “Feature Extraction in Support Vector Machine: A Comparison of PCA, KPCA and ICA,” *Proc. ICONIP '02*, Vol. 2, pp. 1001-1005, Nov. 2002.
- [21] R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, 2nd ed., John Wiley & Sons, 2001.
- [22] D.F. Specht, “Probabilistic neural networks and the polynomial adaline as complementary techniques for classification,” *IEEE Trans. Neural Networks*, Vol. 1, No. 1, pp.111-121, 1990.
- [23] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, 1995.
- [24] <http://s20c.smb.man.ac.uk/services/MIS/MIASmini.html>.
- [25] C.Y. Enderwick and Ecangelia Micheli-Tzanakou, “Classification of mammographic tissue using shape and texture features,” *Proc. IEEE/EMBS*, Vol. 2, pp. 810-813, 1997.