ISTANBUL UNIVERSITY –
JOURNAL OF ELECTRICAL & ELECTRONICS ENGINEERING

YEAR : 2009 (867-875)
VOLUME : 9
NUMBER : 1

MAMMOGRAPHIC MASS CLASSIFICATION USING WAVELET BASED SUPPORT VECTOR MACHINE

Pelin GORGEL[1] Ahmet SERTBAŞ[1] Niyazi KILIC[2]
Osman N. UCAN[2] Onur OSMAN[3]

[1]Istanbul University, Faculty of Engineering, Computer Science Dept. 34320, Avcilar, Istanbul, Turkey
[2]Istanbul University, Faculty of Engineering, Electrical &Electronics Dept. 34320, Avcilar, Istanbul, Turkey
[3]Istanbul Commerce University, Technologies & Programming Dept., Ragip Gumuspala Cad. No.80 Eminonu, Istanbul, Turkey
paras@istanbul.edu.tr

ABSTRACT

In this paper, we investigate an approach for classification of mammographic masses as benign or malign. This study relies on a combination of Support Vector Machine (SVM) and wavelet-based subband image decomposition. Decision making was performed in two stages as feature extraction by computing the wavelet coefficients and classification using the classifier trained on the extracted features. SVM, a learning machine based on statistical learning theory, was trained through supervised learning to classify masses. The research involved 66 digitized mammographic images. The masses were segmented manually by radiologists, prior to introduction to the classification system. Preliminary test on mammogram showed over 84.8% classification accuracy by using the SVM with Radial Basis Function (RBF) kernel. Also confusion matrix, accuracy, sensitivity and specificity analysis with different kernel types were used to show the classification performance of SVM.

Keywords: *Discrete Wavelet Transform, Mammographic Mass Classification, Support Vector Machine, k-fold Cross Validation.*

1. INTRODUCTION

Breast cancer is the most prevalent cancer among women, and it is the leading cause of death from cancer among women of ages 15-54 [1-2]. It has a high mortality but early detection is a key ingredient to reduce this rate. The most effective method for detection of early breast cancer is mammography, which is the most reliable method for the detection of early breast cancer [3] of all diagnostic methods currently available for this purpose. Digital mammography provides additional features not available with standard film-screen mammographic imaging such as contrast enhancement, digital archiving, and computer-aided diagnosis. Unfortunately lesions that are visible on mammograms in retrospect may be missed by radiologists. But previous works on mammograms [4-5] have shown that computer aided detection (CAD) and diagnosis can significantly improve radiologists' accuracy in detecting clustered microcalcifications. CAD systems are designed to provide a second opinion, to aid rather than replace the radiologist. Malignant breast masses often infiltrate the surrounding tissue as they expand. They

Received Date: 27.09.2008
Accepted Date: 05.01.2009

distinguish from benign masses in shape and density [6].

In this study, classification of suspicious areas including benign or malign tumours is realized using update image processing techniques. First we have extracted essential features from mammographic images using discrete wavelet transform [7]. Wavelet technology is a tool for time–space–frequency analysis. Unlike Fourier transform [8], which provides only frequency analysis of signals, wavelet transforms provide time–frequency analysis, which is particularly useful for pattern recognition. After extraction process, we have used machine learning algorithm, Support Vector Machine (SVM) to classify of images under two categories, either malign or benign. As seen in Figure 1, the segmented mammographic images are wavelet decomposed into multi-level low- and high-pass sub-bands, which are then applied as an input of SVMs for training and testing purposes. SVM minimizes structural risk [9] in learning stage. It aims to decrease generalization error instead of directly minimizing learning error. As a result, SVM is able to perform well when applied to data outside the training set. In recent years SVM learning has been used in a wide range of real world applications where it has been found to offer superior performance to that of competing methods [10-11].

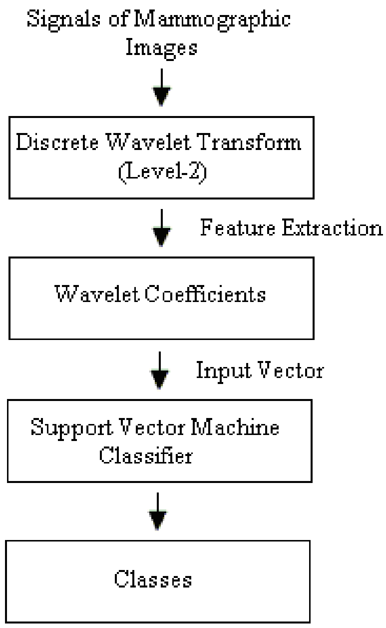


Figure 1. General Structure of the SVM Classifier.

2. MAMMOGRAM DATABASE

In this study, a set of 66 digitized mammograms with 65x65 pixels was used. These images were acquired from Akdeniz University Faculty of Medicine using Giotto Image Md machine with technical properties of 31kV 48mAs Mo/RhLF Grid. 28 patients of 66 have benign breast masses and others have malign masses. As seen in Figure 2, the segments of images consist of the original greyscale image and benign tumours generally present regular and well-defined contours on images, while malign ones usually infiltrate adjacent tissues thus producing irregular and angled edges.

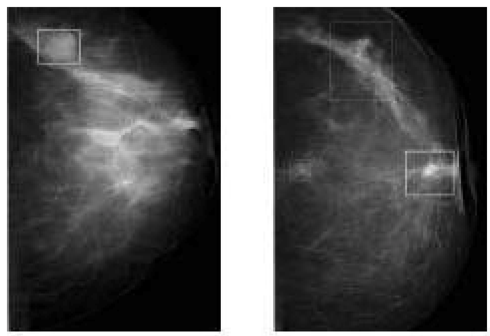


Figure 2. A Sample Mammographic Image from Our Data Set.
a) Benign Breast Mass **b)** Malign Breast Mass

3. FEATURE EXTRACTION

Future extraction is the determination of a feature vector from a pattern. For pattern processing problems to be tractable requires the conversion of patterns to features that are abridged representation of patterns, ideally including only main information. The conversion is realized by Discrete Wavelet Transform for this study.

3.1. Wavelet Decomposition

The wavelet transform is a useful mathematical tool that currently has received a great attention in different applications like compression and denoising of data [12-13]. Unlike the Fourier Transform that gives only information about the frequency components of a signal without specifying the time at which these frequency components occur, the wavelet transform gives information about the frequency and time components of the signal representation simultaneously [14].

Wavelets decompose an image into orthogonal subbands with low-low (LL), low-high (LH), high-low (HL), and high-high (HH) components. The LL subband is further decomposed into another four subbands, and the Low Low Low Low (LLLL) from this second decomposition subband is decomposed once again and so on, as seen in Figure 3. Because of the nature of mammogram few levels of decompositions are necessary to analyze them. In this application two-level decomposition was used.

There are several types of wavelet transforms that can be chosen depending on the application. The continuous wavelet transform (CWT) can be used for continuous signals. In this case, both time and scale are continuous. The discrete wavelet transform (DWT) can also be defined for discrete signals. In this work, the discrete wavelet transform (DWT) was used. The DWT which uses a discrete set of the wavelet scales and translation obeying some defined rules as an implementation of the wavelet transform.
CWT of a function f is:

$$f(a,b) = \int f(x)\psi_{a,b}(x)dx \qquad (1)$$

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \qquad (2)$$

The mother wavelet function is $\psi(x)$. In wavelet transform, the basis functions are derived from scaling and translation of a single function, called mother wavelet. The basis of wavelet function is obtained by scaling and shifting a signal mother wavelet function. $w(x)$ signal is decomposed into a family of synthesis wavelets as given below in equation (3).

$$w(x) = \sum_a \sum_b \left\langle w(x), \psi_{a,b}(x) \right\rangle \psi_{a,b}(x) \qquad (3)$$

$$w[b] = \sum_{i=1 to I} \sum_{k\in Z} c_{i,k}\, g\left[b-2^i k\right] + \sum_{k\in Z} d_{I,k} h_I \left[b-2^I k\right] \qquad (4)$$

$w[b]$ is a discrete time signal, $c_{i,k}$ where $i = 1...I$ are wavelet coefficients and $d_{i,k}$ where $i = 1...I$ are scaling coefficients.

$$c_{i,k} = \sum_b w[b] g_i^*\left[b-2^i k\right],$$
$$d_{i,k} = \sum_b w[b] h_I^*\left[b-2^I k\right] \qquad (5)$$

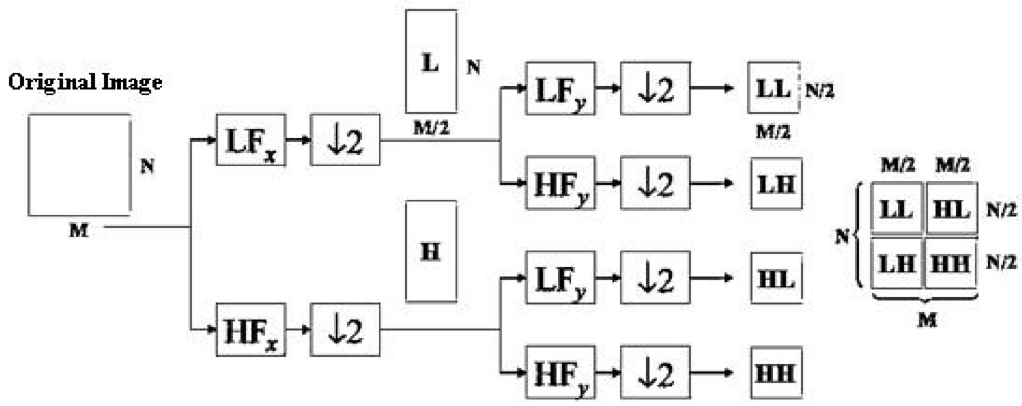


Figure 3. Image Decomposition with Wavelet Transform.

3.2. Discrete Wavelet Transform in Two Dimensions

In this paper a feature vectors is extracted from mammograms based on multilevel wavelets decomposition. These vectors are used to train a SVM for classification of mammograms. The Discrete Wavelet Transform (DWT) is applied to each dimension separately [15]. This yield a multiresolution decomposition of the signal into four subbands called the approximation (low frequency component) and details (high frequency component). The approximation a indicates a low resolution of the original image.

The detail coefficients are horizontal (h), vertical (v) and diagonal (d). An image Y being decomposed into a first level approximation component Y_a^1, and detailed components Y_h^1 , Y_v^1 and Y_d^1 is seen in Figure 4 [16]. The approximation component Y_a contains low frequency components of the image while the detailed components Y_h, Y_v, Y_d contain high frequency components. Thus

$$Y = Y_a^1 + \left\{ Y_h^1 + Y_v^1 + Y_d^1 \right\} \tag{6}$$

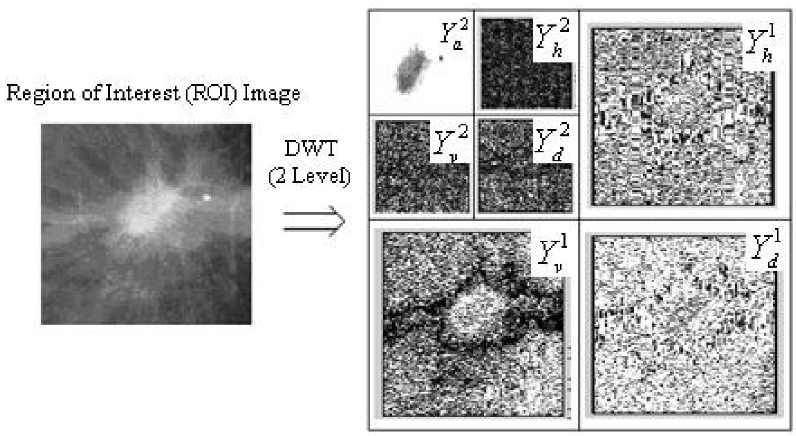
If we apply DWT to Y_a^1 , the second level approximation and detailed components are obtained. If the process is repeated up to N levels, the image Y can be written in terms of the N th approximation component Y_a^N and all detailed components as given below:

$$Y = Y_a^N + \sum\nolimits_{i=1toN} \left\{ Y_h^1 + Y_v^1 + Y_d^1 \right\} \tag{7}$$

At each decomposition level, the half band filters produce signals spanning only half the frequency band. This doubles the frequency resolution as the uncertainty in frequency is reduced by half. The length of the decomposed signals is half the length of the signal in the previous stage. Thus, the first level decomposition of an $N \times N$ image is $N/2 \times N/2$ and the second level decomposition is $N/4 \times N/4$ and so on. If we increase the level of decomposition, compact but coarser approximation of the image is obtained. It is reached that wavelets provide a simple hierarchical framework for interpreting the image information [17].

4. CLASSIFICATION AND TESTING

Recently Support Vector Machine (SVM) which was developed by Vapnik [18] has been used in a range of problems including pattern recognition, bioinformatics and text categorization. SVM provides a novel approach to the two-class classification problem [19]. The use of classifier system in medical diagnosis is increasing gradually and gaining popularity due to many attractive features, and promising empirical performance.



Pelin GORGEL, Ahmet SERTBAŞ, Niyazi KILIC, Osman N. UCAN, Onur OSMAN

Figure 4. Two Level Decomposition of Mammographic Image with DWT.

4.1 SVM Classifier

SVM is a learning tool based on modern statistical learning theory [20]. It gives some useful bounds on the generalization capacity of machines for learning tasks. SVM algorithm constructs a separating hypersurface in the input space by mapping the input space into a high dimensional features space through some non linear mapping chosen a priori (kernel) or constructing in this features space the Maximal Margin Hyperplane [21]. As seen in Figure 5, each side of the hyperplane that separates the data. The separating hyperplane is the hyperplane that maximises the distance between the two parallel hyperplanes. An assumption is made that the larger the margin or distance between these parallel hyperplanes the better the generalisation error of the classifier will be.

When training data $(x_i, y_i), i = 1,...,l$ are separated by $w \cdot x + b = 0$ hyperplane, it happens that $y_i(w \cdot x_i + b) \geq 1$, where $y_i = \pm 1$ are the labels. The margin is $2 / \|w\|$, so the hyperplane, separating data with maximal margin is:

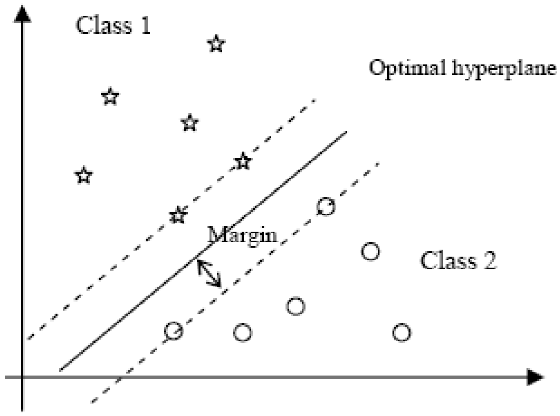


Figure 5. The Separation of Two Classes (indicated by data points marked by “O”s and “* ”s) by an Optimal Hyperplane.

Two parallel hyperplanes are constructed on

$$\begin{cases} \text{minimize} \|w\|^2 / 2 \\ \text{with } y_i(w \cdot x_i + b) \geq 1 \end{cases} \quad (8)$$

Constraints are relaxed to $y_i(w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$ in order to allow misclassification errors. (1) becomes then:

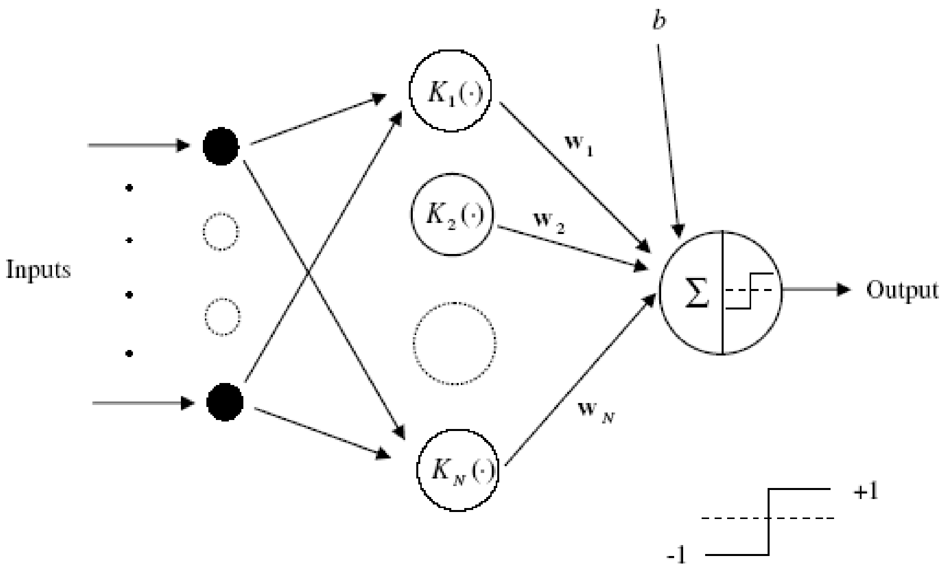


Figure 6. Architecture of the Support Vector Machine (N is the number of support vectors).

4.2 Design of SVM Classifier for Classification

4.2.1 SVM Kernel Functions

SVM is not able to achieve the classification tasks in the nonlinear case. To overcome this limitation on SVM, kernel approaches are developed. The kernel function in an SVM plays the central role of implicitly mapping the input vector into a high-dimensional feature space. Typical choices for kernel function are: Gaussian Radial Basis Function (RBF), Polynomial, Sigmoidal, Inverse multiquadratic, etc. The Gaussian RBF and polynomial kernels are applied in this study.

Polynomial Kernel:

$$K(x_i - x_j) = (x_i^T x_j + 1)^p \qquad (11)$$

where $p > 0$ is a constant.

Gaussian RBF Kernel:

$$K(x, x_i) = \exp[\gamma |x - x_i|] \qquad (12)$$

Both of these kernels satisfy the conditions, and are among the most commonly used in SVM.

4.2.2 Training Examples and SVM Model Selection

A binary support vector machine separates positive samples from the negative samples in training. Lagrange multipliers α_i parameters of each binary support vector machine are determined by minimizing the following cost function,

$$P(w) = \frac{1}{2}\|w\|^2 \qquad (13)$$

Subject to

$$y_i \cdot f(x_i) \geq +1 - \varepsilon_i \;\; \varepsilon_i \geq 0,\; i = 1,2,...k \qquad (14)$$

The cost function L_D is convex and quadratic in terms of the unknown parameters α_i. Here, objective is to maximize the margin of classification subject to constraints. This problem can be solved from dual form which is expressed as

$$L_D = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j \qquad (15)$$

Subject to

$$0 \leq \alpha \leq C \text{ and } \sum_i \alpha_i y_i = 0 \qquad (16)$$

If the value of C regularization parameter that controls the tolerance to classification errors in training is high, then more punishment will be given to the errors. The training vector x_i whose corresponding α_i is nonzero is called support vector.

In this study K-fold cross validation is used for model selection with different criteria, each of which is a prediction of generalization error. *K*-fold cross validation is one way to improve over the holdout method [19]. The samples are partitioned into *K* subsamples (folds). *K*-1 folds are used for training and the last fold is used for evaluation. Cross-validation process is then repeated *K* times leaving one different fold is used for evaluation each time. The advantage of this method is that it matters less how the data gets divided. Every data point gets to be in a test set exactly once, and gets to be in a training set *k-1* times. The variance of the resulting estimate is reduced as *k* is increased [22].

Here, five-fold-cross validation procedure is considered for model selection. Polynomial (linear and quadratic) and RBF kernel are used in training and testing. Due to the fact that best error level is achieved, kernel parameters are chosen as C=100 and γ =0.1.

5. EXPERIMENTAL RESULTS

We have implemented SVM in MATLAB software package with the inputs being the wavelet coefficients of mammographic images, using the OSU SVM [23] for our classification. The second level DAUB4 wavelet approximation coefficients are calculated and given as input to the SVM classifier. In this paper, we apply a SVM classifier to determine whether it is benign mass or malign mass of mammographic images.

As a mentioned section 4, the use of SVM involves training and testing procedure. Here five-fold-cross validation procedure is considered for training and testing with particular kernel function. The RBF and polynomial (linear and quadratic) kernels were used for training and testing. Classification results of the SVM classifier was displayed by a con fusion matrix. Table 1 demonstrates the confusion matrices showing the classification results for the mammogrophic mass images. It shows higher classification rate than the polynomial kernel. According to confusion matrix, 2 of 38 malign mass were classified incorrectly by the network as benign mass; however 8 of 28 benign mass were classified incorrectly by the network as malign mass.

Table 1. Classification results for the mammogrophic mass images.

Kernel Types of SVM Classifier	Desired Result	Output Result	
		Benign mass	Malign mass
RBF	Benign mass	20	8
	Malign mass	2	36
Linear	Benign mass	17	11
	Malign mass	7	33
Quadratic	Benign mass	15	13
	Malign mass	9	29

The test performance of the SVM classifiers can be determined by the computation of sensitivity, specificity and total classification accuracy. These are defined as follows:

Sensitivity: number of correct classified malign mass / number of total malign mass
Specificity: number of correct classified benign mass / number of total benign mass
Total classification accuracy: number of classified mass / number of total mass

Values of sensitivity and specificity and total classification accuracy obtained by using the SVM classifier for classification of mammographic mass images are given in Table 2.

Table 2. Values of statistical parameters of SVM classifiers.

Kernel Types of SVM Classifier	Statistical parameters (%)		
	Sensitivity	Specificity	Total classification accuracy
RBF	94.7	71.4	84.8
Linear	86.8	60.7	75.7
Quadratic	76.3	53.5	66.6

According to table 2, the best results of sensitivity, specificity and total classification accuracy were found by RBF kernel as 94.7%, 71.4% and 84.8% respectively. We can also conclude that SVM with RBF kernel may be one of the promising methods in the classification of mamagrophic mass.

6. CONCLUSION

This study is focused on the classification of masses as benign or malignant. Our research demonstrated that 84.8 % total classification accuracy can be obtained by using the SVM with RBF kernel trained on the wavelet approximation coefficients of decomposed signals. The extraction of features from mammographic images was realized by using Discrete Wavelet Transform, since previous studies suggested that wavelet-based image analysis techniques could occupy a leading position in digital mammography [24]. Wavelets have the ability to discriminate different frequencies and to preserve signal details at different resolutions. The SVM classifier showed a great performance in classifying as it maps the features to a higher dimensional space. In SVM method, two types of kernel are used. The best accuracy results are obtained using RBF kernel and worst classification accuracy results are obtained using quadratic kernel.

7. REFERENCES

[1] B. Sahiner, H.P. Chan, “Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images”, *IEEE T. Med Imaging*, 15:598-610, 1996.

[2] C.C. Boring, T.S. Squires, “Cancer statistics”, *CA-A Cancer J Clin,* 44:7-26, 1994.

[3] H.C. Zuckerman, “The role of mammograph in the diagnosis of breast cancer, in Breast Cancer, Diagnosis and Treatment”, *I. M. Ariel and J. B. Cleary. Eds. McGraw-Hill, New York*, 1987.

[4] H.P. Chan, K. Doi, “Improvement in radiologists' detection of clustered microcalcifications on mammograms: the potential of computer-aided diagnosis”, *Invest Radiol*, 25:1102-1110,1990.

[5] Wang T C, Karayiannis N B, “Detection of microcalcifications in digital mammograms using wavelets” *IEEE T Med Imaging*, 17:498-509, 1989.

[6] Bruce L M, Adhami R R, “Classifying mammographic mass shapes using the wavelet transform modulus-maxima method”, *IEEE T Med Imaging* 18:1170-1177, 1999.

[7] S.G. Mallat, “A theory of multiresolution signal decomposition: the wavelet representation”, *IEEE T. Pattern Anal*, 1980, 11:674–693.

[8] R.N. Bracewell, “The Fourier Transform and its Applications”, *McGraw-Hill, New York*, 1999.

[9] V. Vapnik, “Statistical learning theory”, *Wiley, New York*, 1998.

[10] K.R. Muller, S. Mika, “An introduction to kernel-based learning algorithms”, *IEEE T., Neural Network*, 2001, 12:181-201.

[11] I. El-Naqa, Y. Yang, “A support vector machine approach for detection of microcalcifications in mammogram”, *IEEE T. Med Imaging*, 2002, 21:1552-1563.

[12] C.M. Brislawn, “Fingerprints go digital”, *Notices Amer Math Soc*., 1995, 42:1278-1283.

[13] M. Al-qdah, A.R. Ramli, “Detection of calcifications in mammography using wavelets”, *Student Conference on Research and Development (SCORcD) Proceedings*, 2003, Putrajaya, Malaysia.

[14] Mathsoft Wavefet resources, *A great collection of theory and application oriented articles on the web at* http://mw.mthsof.cod wavelets.html, 1997.

[15] S. Chaplot, L.M. Patnaik, “Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network”, *Biomedical Signal Processing and Control*, 2006, 1:86-92.

[16] R.C. Gonzalez, R.E. Woods, “Digital Image Processing”, *Prentice Hall, New Jersey*, 2002.

[17] Koenderink J (1984) The structure of images. Biol Cybern 50:363-370.

[18] V.N. Vapnik, “An overview of statistical learning theory”, *IEEE T. Neural Network*, 1999 10:988-999.

[19] K. Polat, S. Guneş, “Breast cancer diagnosis using least square support vector machine”, *Digital Signal Process.* “in press”, 2006.

[20] E.D. Übeyli, “ECG beats classification using multiclass support vector machines with error correcting output codes”, *Digital Signal Process.,*17:675-684, 2007.

[21] A. Bazzani, A.D. Bevilacqua, “Automatic detection of clustered microcalcifications in digital mammograms using an SVM classifier”, ESANN'2000 proceedings, European Symposium on Artificial Neural Networks, Bruges, Belgium, 2000.

[22] P.A. Devijver, J. Kittler, “Pattern Recognition: A Statistical Approach”, *Prentice-Hall, London*, 1982.

[23] C.C. Chang, C.J. Lin, LIBSVM:a library for support vector machines, Software available athttp://www.csie.ntu.edu.tw/~cjilin/libsvm2001.

[24] A.F. Laine, S., "Schuler Mammographic feature enhancement by multiscale analysis", *IEEE T. Med Imaging*, 13:725-740,1994.