
Support Vector Machines for Diagnosis of Breast Tumors on US Images¹

Ruey-Feng Chang, PhD, Wen-Jie Wu, MS, Woo Kyung Moon, MD, Yi-Hong Chou, MD, Dar-Ren Chen, MD

Rationale and Objectives. Breast cancer has become the leading cause of cancer deaths among women in developed countries. To decrease the related mortality, disease must be treated as early as possible, but it is hard to detect and diagnose tumors at an early stage. A well-designed computer-aided diagnostic system can help physicians avoid misdiagnosis and avoid unnecessary biopsy without missing cancers. In this study, the authors tested one such system to determine its effectiveness.

Materials and Methods. Many computer-aided diagnostic systems for ultrasonography are based on the neural network model and classify breast tumors according to texture features. The authors tested a refinement of this model, an advanced support vector machine (SVM), in 250 cases of pathologically proved breast tumors (140 benign and 110 malignant), and compared its performance with that of a multilayer propagation neural network.

Results. The accuracy of the SVM for classifying malignancies was 85.6% (214 of 250); the sensitivity, 95.45% (105 of 110); the specificity, 77.86% (109 of 140); the positive predictive value, 77.21% (105 of 136); and the negative predictive value, 95.61% (109 of 114).

Conclusion. The SVM proved helpful in the imaging diagnosis of breast cancer. The classification ability of the SVM is nearly equal to that of the neural network model, and the SVM has a much shorter training time (1 vs 189 seconds). Given the increasing size and complexity of data sets, the SVM is therefore preferable for computer-aided diagnosis.

Key Words. Breast neoplasms, diagnosis; breast neoplasms, US; computers, diagnostic aid.

© AUR, 2003

Recent statistics show that breast cancer affects one of every 10 women in Europe and one of every eight in the United States (1). Early diagnosis and early treatment are the best ways to reduce deaths due to breast cancer. Early diagnosis requires an accurate and reliable diagnostic procedure that allows physicians to distinguish benign from

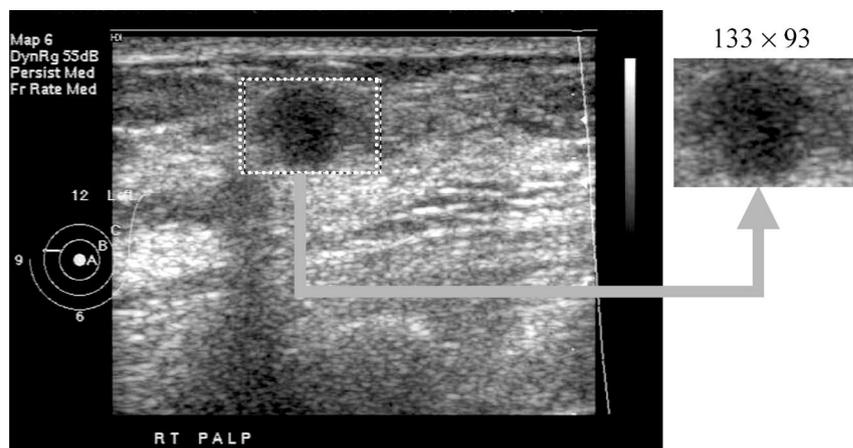
malignant breast tumors, and finding such a procedure is an important goal. Current procedures for early detection and diagnosis of breast cancer include self-examination, mammography (2–4), and ultrasonography (US) (5). Previously, the recommended role of US was limited to differentiating between cysts and solid masses, evaluating masses in a dense breast, and guiding interventional procedures. Important technical advances have been made recently in diagnostic US with the introduction of higher-frequency linear transducers. The increased computing power of US platforms has enabled the production of fully digital systems with improved resolution and image contrast, and this digitization in turn greatly helps in image processing. Hence, the role of breast US has expanded beyond the distinction between cystic and solid lesions. US is a convenient and safe diagnostic method,

Acad Radiol 2003; 10:189–197

¹ From the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan (R.F.C., W.J.W.); the Department of Diagnostic Radiology, Seoul National University Hospital, Seoul, South Korea (W.K.M.); the Department of Radiology, Veterans General Hospital, Taipei, Taiwan (Y.H.C.); and the Department of General Surgery, China Medical College and Hospital, 2 Yer-Der Rd, Taichung, Taiwan (D.R.C.). Received September 3, 2002; revision requested September 24; revision received and accepted October 10. **Address correspondence to** D.R.C.

© AUR, 2003

Figure 1. A 640×480 digital image obtained by using a US scanner. In a rectangle 1×1 cm, there are $100 \times 100 = 10,000$ pixels. The ROI rectangle is 1.33×0.93 cm and 133×93 pixels.



but there is overlap between benignancy and malignancy in US appearances, and interpretation is subjective. In 1995, Stavros et al (6) showed that the sensitivity of breast US for malignancy was 98.4%; the specificity, 67.8%; and the overall accuracy, 72.9%. These results were achieved by experienced radiologists.

The support vector machine (SVM) (7–9) has become an effective tool for pattern recognition, machine learning, and data mining, because of its high generalization performance. Given a set of points that all belong to one of two classes, an SVM can find the hyperplane that leaves the largest possible fraction of points of the same class on the same side, while maximizing the distance of either class from the hyperplane. This *optimal separating hyperplane* can minimize the risk of misclassifying examples of the test set.

Neural networks (10) also have been successful in many applications, especially for clustering (11) and pattern recognition (12). Recent research, however, has suggested that the SVM is superior to the neural network (13–15). To verify this, and to test our method, we used a multilayer propagation (MLP) neural network model, as well as an SVM, to classify tumors in our experiments.

We evaluated breast masses in a series of pathologically proved tumors by using autocovariance texture parameters and the SVM to classify the tumors in regions of interest (ROIs) of US images selected by the physician. In this article, we describe how we acquired the experimental image data and extracted features used to classify tumors. We also present our method for using the SVM to distinguish benign from malignant breast tumors, compare it with a diagnostic method based on an MLP neural network model, and present and discuss our results.

IMAGE DATA ACQUISITION

Our US image database includes 250 US images of pathologically proved benign breast tumors from 140 patients, and carcinomas from 110 patients; the diagnoses were proved by means of fine-needle cytology, core-needle biopsy, or open biopsy. All US images were obtained by a surgeon (D.R.C.), who also selected the ROIs. Data were consecutively collected from August 1, 1999, to May 31, 2000, and patients' ages ranged from 18 to 64 years. Only one image from each patient was included in the database.

The US images were obtained by using an ATL HDI 3000 system (Philips Medical Systems, Bothell, Wash) with an L10-5 small-parts transducer, a linear-array transducer with a frequency of 5–10 MHz and a scan width of 38 mm. Dynamic range and mapping were set at 55 decibels and at 6 decibels, respectively. During image acquisition, patients were supine with arms extended overhead. No acoustic standoff pad was used.

The monochrome US images were quantized into 8 bits (ie, 256 gray levels) and their features were stored on magneto-optical disks. These images could be read and analyzed on a personal computer and served as our experimental data. W.K.M. supplied the database. The subimage of the ROI was manually selected by D.R.C. (who was unfamiliar with the tissue diagnosis and cell-type ROI selections), using the ProImage package (Prolab, Taipei, Taiwan). The ROI subimage was then saved as a file for later analysis. Figure 1 is an example of a real-time digitized monochromatic US image of a benign tumor. ROI subimages were used in our database of breast im-

Table 1
Mean Autocovariance Coefficients for Benign and Malignant Cases

Autocovariance Coefficient	Benign Cases*	Malignant Cases*	Mean Difference
A(1,0)	0.967 ± 0.011	0.972 ± 0.007	0.005
A(2,0)	0.902 ± 0.029	0.917 ± 0.019	0.015
A(3,0)	0.834 ± 0.049	0.858 ± 0.032	0.024
A(4,0)	0.778 ± 0.064	0.809 ± 0.042	0.031
A(0,1)	0.893 ± 0.027	0.926 ± 0.015	0.033
A(1,1)	0.870 ± 0.032	0.907 ± 0.018	0.036
A(2,1)	0.823 ± 0.043	0.866 ± 0.026	0.043
A(3,1)	0.771 ± 0.056	0.821 ± 0.035	0.050
A(4,1)	0.728 ± 0.067	0.783 ± 0.043	0.055
A(0,2)	0.734 ± 0.060	0.812 ± 0.035	0.078
A(1,2)	0.723 ± 0.061	0.803 ± 0.036	0.080
A(2,2)	0.700 ± 0.066	0.783 ± 0.039	0.082
A(3,2)	0.674 ± 0.071	0.759 ± 0.043	0.085
A(4,2)	0.649 ± 0.077	0.737 ± 0.047	0.088
A(0,3)	0.617 ± 0.082	0.737 ± 0.048	0.120
A(1,3)	0.611 ± 0.082	0.732 ± 0.048	0.120
A(2,3)	0.600 ± 0.084	0.722 ± 0.049	0.122
A(3,3)	0.586 ± 0.086	0.709 ± 0.051	0.123
A(4,3)	0.571 ± 0.089	0.695 ± 0.053	0.124
A(0,4)	0.534 ± 0.095	0.690 ± 0.058	0.156
A(1,4)	0.529 ± 0.095	0.686 ± 0.057	0.157
A(2,4)	0.521 ± 0.095	0.678 ± 0.058	0.157
A(3,4)	0.511 ± 0.096	0.669 ± 0.058	0.158
A(4,4)	0.499 ± 0.098	0.658 ± 0.059	0.158

*Numbers are means ± standard deviations.

ages to investigate the texture characteristics of benign and malignant tumors.

FEATURE EXTRACTION

Many texture features have been proposed for texture analysis, and they can be classified in three main groups: models (16,17), mathematical morphology (18), and statistical methods (19). In this section, we briefly introduce autocorrelation and autocovariance coefficients. Autocorrelation coefficients are derived from statistical analysis. The advantage of these texture features is that they reflect the correlation among pixels within an image. Following the example of Gonzalez and Woods (20), we define the normalized autocorrelation coefficient $\gamma(\Delta m, \Delta n)$ between pixel (i, j) and pixel $(i + \Delta m, j + \Delta n)$ in an image with size $M \times N$ as follows:

$$\gamma(\Delta m, \Delta n) = \frac{A(\Delta m, \Delta n)}{A(0,0)}, \quad (1)$$

where

$$A(\Delta m, \Delta n) = \frac{1}{(M - \Delta m)(N - \Delta n)} \cdot \sum_{x=0}^{M-1-\Delta m} \sum_{y=0}^{N-1-\Delta n} f(x, y)f(x + \Delta m, y + \Delta n). \quad (2)$$

The autocorrelation method has a disadvantage, however. It is usually affected by brightness. If two images have similar textures but different brightness, their autocorrelation coefficients may be different. Thus, we replace autocorrelation coefficients with autocovariance coefficients as our image features. The autocovariance method is defined as follows:

$$A(\Delta m, \Delta n) = \frac{1}{(M - \Delta m)(N - \Delta n)} \cdot \sum_{x=0}^{M-1-\Delta m} \sum_{y=0}^{N-1-\Delta n} [f(x, y) - \bar{f}][f(x + \Delta m, y + \Delta n) - \bar{f}], \quad (3)$$

where \bar{f} is the mean value of $f(x, y)$.

We used this method to characterize the texture features of each US image, representing each image initially by a 5×5 autocovariance matrix (ie, 25 autocovariance coefficients; both Δm and $\Delta n = 5$). We decided to discard the $\gamma(0,0)$ coefficient, however, because the value of $\gamma(0,0)$ is always 1 for the normalized autocovariance matrix. We combined the other 24 autocovariance coefficients to form a 24-dimensional image features vector.

In Table 1, we list the means and standard deviations of the autocovariance coefficients for benign and malignant tumors and the mean differences between the two groups. These mean differences are substantial, which indicates that the autocovariance coefficients are useful for distinguishing benign from malignant tumors. Moreover, Figure 2 shows that as Δm and Δn increase, so does the mean difference.

CLASSIFYING TUMORS BY USING THE SVM AND THE MLP NEURAL NETWORK

In this section, we introduce SVMs. While we used an SVM to distinguish benign from malignant tumors, we also compared it with an MLP neural network model to

prove the SVM's superior classification ability. MLP neural networks are discussed below.

SVM Classification

SVMs have been recently proposed as effective for many applications, because of their high generalization performance. Here we will first introduce the simple case of a linearly separable set and then describe the concept of SVMs and expand it to the more general, nonseparable case. Finally, we will introduce the general case of non-linear separating surfaces.

Optimal separating hyperplane.—If we have a training example set $S = \{(x_i, y_i)_{1 \leq i \leq N}\}$, and each example $\mathbf{x}_i \in R^n$ belongs to a class labeled by $y_i \in \{-1, 1\}$, our object is to find a hyperplane that divides S , leaving all the points with the same label on the same side of the hyperplane. Meanwhile, we also maximize the distance between the two classes and the hyperplane. This means we must find a pair (\mathbf{w}, b) such that

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 0, \quad i = 1, \dots, N, \quad (4)$$

where $\mathbf{w} \in R^n$ and $b \in R$. According to the pair (\mathbf{w}, b) , we can achieve an equation of a separating hyperplane (Fig 3),

$$\mathbf{w} \cdot \mathbf{x} + b = 0. \quad (5)$$

We can say that the set S is linearly separable if there is at least one hyperplane satisfying Equation (4). Meanwhile, we can rescale \mathbf{w} and b so that

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, N. \quad (6)$$

The minimal distance between the closest point and the hyperplane is $1/\|\mathbf{w}\|$, and the margin is $2/\|\mathbf{w}\|$. The margin is a measure of generalizability. The larger the margin, the better the generalization.

Among the separating hyperplanes, there must be one from which the distance to the closest point is maximal—the optimal separating hyperplane (OSH)—and which will maximize the margin. The goal of the SVM is to find the OSH of the set S . For the OSH to be found, $\|\mathbf{w}\|^2$ must be minimized under constraint Equation (6).

According to the property that $\|\mathbf{w}\|^2$ is convex (7–9), we can minimize it under constraint Equation (6) by means of the classic method of Lagrange multipliers. Hence, if $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)$ is the N nonnegative La-

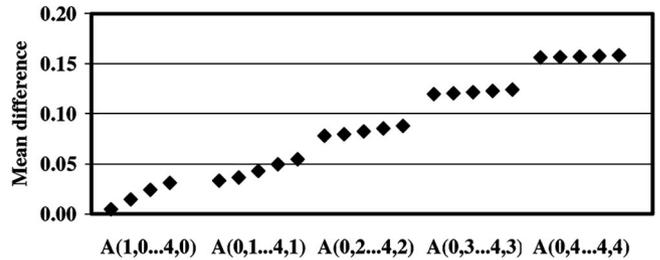


Figure 2. Mean differences between autocovariance coefficients for benign and malignant cases.

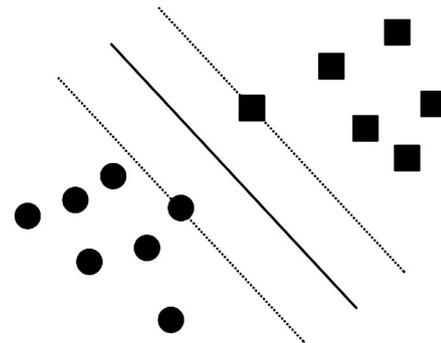


Figure 3. Separating hyperplane (dashed lines identify the margin).

grange multipliers associated with constraint Equation (6), the problem of finding the OSH is equivalent to the maximization of the function

$$W(\boldsymbol{\alpha}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j, \quad (7)$$

where $\alpha_i \geq 0$ and under constraint

$$\sum_{i=1}^N y_i \alpha_i = 0.$$

If we denote the vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)$ as the solution of Equation (7), then the OSH (\mathbf{w}, b) has the following expansion:

$$\bar{\mathbf{w}} = \sum_{i=1}^N \bar{\alpha}_i y_i \mathbf{x}_i, \quad (8)$$

while b can be determined from $\boldsymbol{\alpha}$ and from the Kuhn-Tucker conditions (21), as follows:

$$\bar{\alpha}_i[y_i(\bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b}) - 1] = 0, \quad i = 1, 2, \dots, N. \quad (9)$$

If the training examples (\mathbf{x}_i, y_i) correspond with nonzero coefficients α_i , then we call them *support vectors*. Finally, the decision function of classifying a new data point \mathbf{x} can be written as follows:

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^N \bar{\alpha}_i y_i \mathbf{x}_i \cdot \mathbf{x} + \bar{b}\right). \quad (10)$$

Linearly nonseparable case.—In the preceding section, we supposed that the set S was linearly separable and we introduced the concept of the SVM. Now, we expand set S to the linearly nonseparable set S' . Since the set S' is linearly nonseparable, we must introduce N nonnegative variables $\xi = (\xi_1, \xi_2, \dots, \xi_N)$, such that

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N. \quad (11)$$

We call these *slack variables*. Their purpose is to allow misclassified points corresponding with $\xi_i > 1$. Hence, the generalized OSH is the solution of the following minimizing problem,

$$\frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^N \xi_i, \quad (12)$$

where C is a regularization parameter. If the parameter C is small, the OSH tends to maximize the distance $1/\|\mathbf{w}\|$, while a larger C will cause the OSH to minimize the number of misclassified points.

Nonlinear SVMs.—Most training sets that we want to classify are linearly nonseparable. We may be able to solve these problems by introducing slack variables. Even if we do, however, the classification results will not be optimal. Instead of using slack variables, we can transfer data from the original low-dimensional feature space into a high-dimensional one. Through this transformation, the OSH can be constructed more easily, and we can achieve a classifier with better generalization.

Let $\Phi(\mathbf{x})$ denote a mapping function that maps \mathbf{x} into a high-dimensional feature space. We can then rewrite Equation (7) as follows:

$$W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j). \quad (13)$$

Now, let $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$. Equation (13) can then be rewritten as

$$W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j), \quad (14)$$

where K is called a *kernel function* and must satisfy the Mercer theorem. Finally, we can achieve a new decision function, as follows:

$$f(\mathbf{x}) = \text{sgn}\left[\sum_{i=1}^N \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right]. \quad (15)$$

Several common kernel functions are used to map data into high-dimensional feature space:

Linear kernel:

$$K(\mathbf{x}, \mathbf{z}) = \mathbf{x} \cdot \mathbf{z}. \quad (16)$$

Polynomial kernel:

$$K(\mathbf{x}, \mathbf{z}) = (\gamma \cdot \mathbf{x} \cdot \mathbf{z} + \text{coef})^d, \quad (17)$$

where γ and coef = constants and d = a degree.

Gaussian radial basis kernel:

$$K(x, z) = \exp(-\gamma \cdot |\mathbf{x} - \mathbf{z}|^2), \quad (18)$$

where γ is a constant.

Sigmoidal neural network kernel:

$$K(\mathbf{x}, \mathbf{z}) = \tanh(\gamma \cdot \mathbf{x} \cdot \mathbf{z} + \text{coef}), \quad (19)$$

where γ and coef are constants.

In this study, we mainly used a nonlinear SVM with a Gaussian radial basis kernel as our classifier. Autocovariance textures are used as inputs to find an OSH for distinguishing benign tumors from malignant ones.

MLP Neural Network

Neural networks also have been used successfully in many applications. They consist of many simple computing units, called *neurons* or *processing units*, and massive interconnections between these units. Through these neurons and interconnections, neural networks can store experiential knowledge and make it available for use (22).

Table 2
Classification of Breast Tumors with the SVM
and the MLP Neural Network

Sonographic Classification	SVM		MLP	
	Benign	Malignant	Benign	Malignant
Benign	109 TN	5 FN	108 TN	6 FN
Malignant	31 FP	105 TP	32 FP	104 TP
Total	140	110	140	110

Note.—“Benign” and “Malignant” in the column headings indicate histologic findings. TN = true-negative, FN = false-negative, FP = false-positive, TP = true-positive.

In the Appendix, we briefly introduce the basic components of a neural network and describe the architecture and learning algorithm of an MLP neural network.

In this study, we used an MLP neural network with 25 input nodes, 10 hidden nodes, and a single output node to classify tumors. The autocovariance textures (24 dimensions) and a predefined terminal error threshold (τ) of input layer are used as the input data of the MLP neural network. The experimental results obtained with the MLP neural network were used as a reference and compared with those obtained with the SVM.

SIMULATIONS AND RESULTS

Our experiments were conducted with a database of 250 pathologically proved cases (140 benign and 110 malignant breast tumors). These sonographic images were randomly divided into five groups. We set the first group as a testing group and used the remaining four groups to train the SVM. After training, the SVM was then tested on the first group. We then set the second group as a testing group. This process was repeated until each of the five groups was set as a testing group.

We used a nonlinear SVM with a Gaussian radial basis kernel as our classifier, where $C = 975$ and $\gamma = 0.061$. To compare the performance of our proposed method, we used an MLP neural network classifier. We set its learning rate η at 0.02, the threshold τ at 0.000001, and the maximal number of iterations at 20,000. The simulations were performed by using an Intel Pentium-VI 2-GHz personal computer (ASUS, Taipei, Taiwan) with a single central processing unit and a Microsoft Windows XP operating system. Table 2 lists the classification results.

To estimate performance, we used five objective indexes: accuracy, sensitivity, specificity, positive predictive

Table 3
Summary of Performance for the SVM and the MLP
Neural Network

Index	SVM	MLP
Accuracy (%)	85.60	84.80
Sensitivity (%)	95.45	94.55
Specificity (%)	77.86	77.14
Positive predictive value (%)	77.21	76.47
Negative predictive value (%)	95.61	94.74

Note.—Accuracy = $(TP + TN)/(TP + TN + FP + FN)$, sensitivity = $TP/(TP + FN)$, specificity = $TN/(TN + FP)$, positive predictive value = $TP/(TP + FP)$, and negative predictive value = $TN/(TN + FN)$. TP = true-positive, TN = true-negative, FP = false-positive, and FN = false-negative.

value, and negative predictive value. The accuracy of SVM for classifying malignancies was 85.6% (214 of 250); the sensitivity, 95.45% (105 of 110); the specificity, 77.86% (109 of 140); the positive predictive value, 77.21% (105 of 136); and the negative predictive value, 95.61% (109 of 114). Table 3 compares the performance of the SVM with that of the MLP neural network. Figure 4 illustrates the receiver operating characteristic (ROC) curves for the SVM and the MLP neural network in the classification of malignant and benign tumors. The A_z value for the ROC curve is 0.9396 ± 0.0145 (standard deviation) for the SVM and 0.9395 ± 0.0141 for the MLP neural network. The P value of the difference between the areas A_z under the two ROC curves (z test) is .9942. As Table 3 and Figure 4 demonstrate, the classification ability of the SVM is equal to that of the MLP neural network.

To prove that the autocovariance coefficient method of analysis works for other texture features, as well as for global regions, we divided the experimental data into dark and bright images and then compared the classification results (Table 4). The results for dark and bright images were almost the same. Hence, the autocovariance coefficient method is not affected by brightness. Finally, we also compared the training time for the SVM and the MLP neural network and found that the SVM was much less time consuming, with a training time of only 1 second compared with 189 seconds for the MLP neural network.

CONCLUSION

With the rapid development of US technologies in recent years, many different US systems are currently used

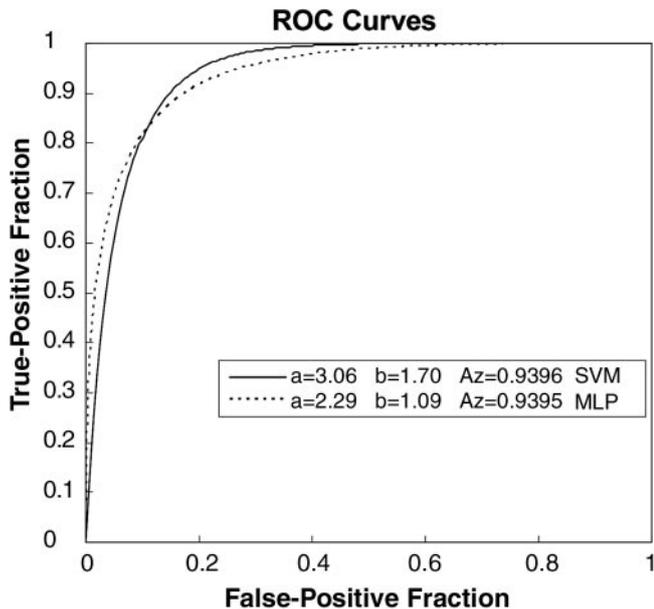


Figure 4. Diagram of ROC curves for the SVM and the MLP neural network in the classification of malignant and benign tumors. A_z value is 0.9396×0.0145 for the SVM and 0.9395×0.0141 for the MLP neural network.

Table 4
Performance of the SVM and the MLP Neural Network for Dark and Bright Images

Index	Dark Images		Bright Images	
	SVM	MLP	SVM	MLP
Accuracy (%)	85.60	82.40	85.60	87.20
Sensitivity (%)	96.36	92.73	94.55	96.36
Specificity (%)	77.14	74.29	78.57	80.00
Positive predictive value (%)	76.81	73.91	77.61	79.10
Negative predictive value (%)	96.43	92.86	94.83	96.55

in diagnosis. Improved diagnostic imaging techniques, in turn, have enabled less-invasive treatment of detected breast lesions. We propose a diagnostic system that uses an SVM in place of the commonly used neural network to differentiate between benign and malignant tumors.

Texture features are helpful for classifying benign and malignant tumors on sonograms. The potential of sonographic texture analysis to improve breast tumor diagnosis has already been demonstrated (23,24). There are three main types of texture analysis: models, mathematical morphology, and statistical methods. Autocovariance texture is a kind of statistical method. In the proposed diagnostic system, autocovariance texture features are used to classify tumors with an SVM model. To demonstrate the

performance of the SVM, we compared its classification ability with that of an MLP neural network model. From our results, we conclude that the SVM performs as well as the MLP neural network. As for training time, the SVM needs only one iteration in training data while the MLP neural network needs many; the SVM is up to 189 times faster. Given the increasing size and complexity of data sets, SVM is more suited to computer-aided diagnosis than MLP neural networks.

APPENDIX

The neuron is the basic computing unit of a neural network. A model of a neural network is shown in Figure A1. It consists of four basic elements: (a) a set of weights: each input signal x_i is multiplied by weight w_i , and p is the number of input signals; (b) the adder, an operation used for summing the weighted signals; (c) an activation function, for limiting the amplitude of the output signal, chosen to satisfy some specification of the problem that the neural network is attempting to solve; and (d) the threshold, a parameter that lowers the input signal of the activation function.

We can formulate the neural network in the following mathematical terms:

$$a = \sum_{i=1}^p w_i x_i$$

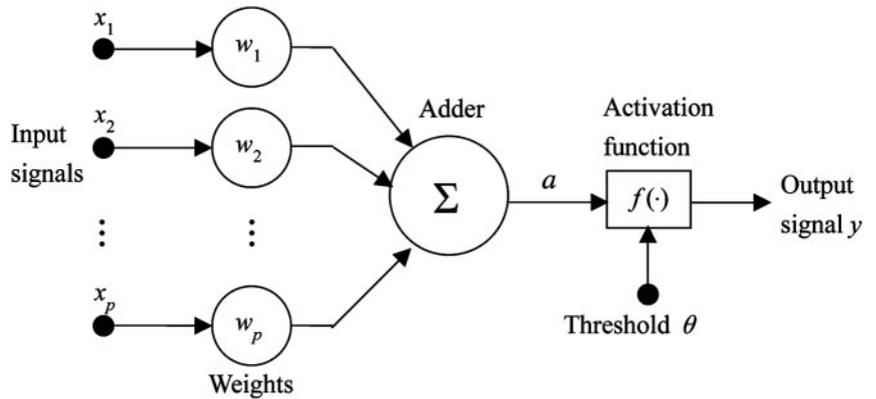
$$y = f(a - \theta), \tag{A1}$$

where x_1, x_2, \dots, x_p = input signals, w_1, w_2, \dots, w_p = weights, a = the output of the adder, θ = the threshold, $f(\cdot)$ = the activation function, and y = the output signal of the neural network.

MLP neural networks are an important class of neural networks. In general, there is at least one hidden layer in an MLP neural network, and the function of its neurons is to arbitrate between the neural network's input and output. The architecture of an MLP neural network is shown in Figure A2.

The backpropagation algorithm, the most popular learning algorithm, is usually used in an MLP neural network. It can typically be divided into two phases: forward and backward. In the forward phase, input signals are propagated forward through the network, and output signals are produced in the output layer. Meanwhile, error

Figure A1. Model of a neural network.



signals are generated by comparing the produced output with the desired response. In the backward phase, error signals are propagated backward through the network, and some parameters, such as weights, can be adjusted in reference to the error signals. The backpropagation algorithm includes four steps:

1. Initialization: Set initial weights and learning rate η for all layers of the MLP neural network. A terminating error threshold value τ is also selected to stop the learning process.
2. Forward computation: Calculate the output values of the MLP neural network layer by layer. We define the internal output signal $h_j^{(l)}(n)$ for neuron j in layer l at iteration n as follows:

$$h_j^{(l)}(n) = \sum_{i=0}^k w_{ji}^{(l)}(n)p_i^{(l-1)}(n), \quad (A2)$$

where $p_i^{(l-1)}(n)$ is the output signal of neuron i in layer $l - 1$, $w_{ji}^{(l)}(n)$ is the weight between neuron j in layer l and neuron i in layer $l - 1$, and k is the dimension of input vectors. The output signal of neuron j is defined as

$$p_j^{(l)}(n) = \begin{cases} x_j(n) & \text{if neuron } j \text{ is in the input layer,} \\ o_j(n) & \text{if neuron } j \text{ is in the output layer,} \\ f(h_j^{(l)}(n)) & \text{otherwise,} \end{cases} \quad (A3)$$

where $x_j(n)$ = the j th element of the input vector $\mathbf{x}(n)$. Then we can achieve the error signal $e_j(n)$, which is defined as

$$e_j(n) = z_j(n) - o_j(n), \quad (A4)$$

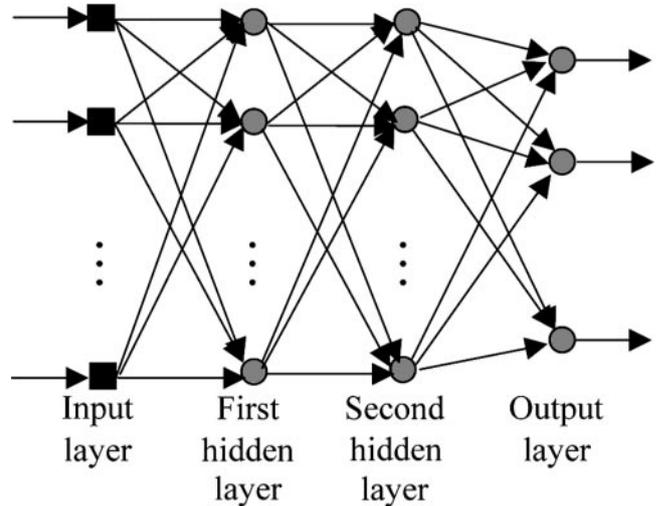


Figure A2. Structural graph of an MLP neural network.

where $z_j(n)$ = the j th element of the desired output vector $\mathbf{z}(n)$ and $o_j(n)$ = the j th element of the produced output $\mathbf{o}(n)$.

3. Backward computation: Calculate the local gradients δ of the MLP neural network layer by layer. The local gradients can point to the required change in respective weight. They are defined as follows:

$$\delta_j^{(l)}(n) = \begin{cases} e_j(n)o_j(n)[1 - o_j(n)] & \text{for neuron } j \text{ in output layer} \\ h_j(n)[1 - h_j(n)] \sum_{i=1}^m \delta_i^{(l)}(n)w_{ij}^{(l+1)}(n) & \text{for neuron } j \text{ in hidden layer,} \end{cases} \quad (A5)$$

where m is the total number of neurons in layer l . Then the connection weights between layer l and layer $l - 1$ are modified according to

$$\Delta w_{ji}^{(l)} = \eta \delta_j^{(l)}(n) h_i^{(l-1)}(n). \quad (\text{A6})$$

4. Iteration of learning procedure: An average distortion function is defined as follows:

$$D = |\text{SE}_{\text{avg}}[\mathbf{w}(n)] - \text{SE}_{\text{avg}}[\mathbf{w}(n-1)]|, \quad (\text{A7})$$

where $\text{SE}_{\text{avg}}[\mathbf{w}(n)]$ is average squared errors for the training samples with the weight vector $\mathbf{w}(n)$ in iteration n . The learning procedure will iteratively execute until the stopping criterion (ie, $D < \tau$) is satisfied.

REFERENCES

- Bothorel S, Meunier BB, Muller SA. Fuzzy logic based approach for semilogical analysis of microcalcification in mammographic images. *Int J Intell Syst* 1997; 12:819–848.
- Dhawan AP, Chitre Y, Kaiser-Bnoasso C, Moskowitz M. Analysis of mammographic microcalcifications using gray-level image structure features. *IEEE Trans Med Imaging* 1996; 15:246–259.
- Petrick N, Chan HP, Sahiner B, Wei D. An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. *IEEE Trans Med Imaging* 1996; 15:59–67.
- Sahiner B, Chan HP, Petrick N, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE Trans Med Imaging* 1996; 15:598–610.
- Shankar PM, Reid JM, Ortega H, Piccoli CW, Goldberg BB. Use of non-Rayleigh statistics for identification of tumors in ultrasonic B-scans of the breast. *IEEE Trans Med Imaging* 1993; 12:687–692.
- Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology* 1995; 196:123–134.
- Pontil M, Verri A. Support vector machines for 3D object recognition. *IEEE Trans Pattern Analysis Machine Intelligence* 1998; 20:637–646.
- Vapnik VN. *The nature of statistical learning theory*. New York, NY: Springer-Verlag, 1995.
- Chapelle O, Haffner P, Vapnik VN. Support vector machines for histogram-based image classification. *IEEE Trans Neural Networks* 1999; 10:1055–1064.
- Hagan MT, Demuth HB, Beale MH. *Neural network design*. Boston, Mass: PWS, 1996.
- Park DC. Centroid neural network for unsupervised competitive learning. *IEEE Trans Neural Networks* 2000; 11:520–528.
- Gader PD, Keller JM, Krishnapuram R, Chiang JH, Mohamed MA. *Neural and fuzzy methods in handwriting recognition*. Computer 1997; 30:79–86.
- Burbidge R, Trotter M, Buxton B, Holden S. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput Chem* 2001; 26:5–14.
- Ding CH, Dubchak I. Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics* 2001; 17:349–358.
- Liang H, Lin Z. Detection of delayed gastric emptying from electrogastragrams with support vector machine. *IEEE Trans Biomed Eng* 2001; 48:601–604.
- Hu R, Fahmy MM. Texture segmentation based on a hierarchical Markov random field model. *Signal Process* 1992; 26:285–305.
- Chen CC, Daponte JS, Fox MD. Fractal feature analysis and classification in medical imaging. *IEEE Trans Med Imaging* 1989; 8:133–142.
- Chen Y, Dougherty ER, Totterman SM, Hornak JP. Classification of trabecular structure in magnetic resonance images based on morphological granulometries. *Magn Reson Med* 1993; 29:358–370.
- Estevez L, Kehtarnavaz N, Wendt R III. Interactive selective and adaptive clustering for detection of microcalcifications in mammograms. *Dig Signal Proc Rev J* 1996; 6:224–232.
- Gonzalez RC, Woods RE. *Image compression*. In: *Digital image processing*. Reading, Mass: Addison-Wesley, 1992; 312–315.
- Kuhn HK, Tucker AW. *Nonlinear programming*. Presented at the Berkeley Symposium on the Mathematics of Statistical Probability, Berkeley, Calif, 1961.
- Aleksander I, Morton H. *An introduction to neural computing*. London, England: Chapman & Hall, 1990.
- Garra BS, Krasner BH, Horii SC, Ascher S, Mun SK, Zeman RK. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. *Ultrason Imaging* 1993; 15: 267–285.
- Goldberg V, Manduca A, Ewert DL, Gisvold JJ, Greenleaf JF. Improvement in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence. *Med Phys* 1992; 19:1475–1481.