Classification of breast masses using a committee machine of artificial neural networks
Classification of breast masses using a committee machine of artificial neural networks

Leandro Augusto da Silva
Emilio Del Moral Hernandez
University of Sao Paulo
Department of Electronic Systems Engineering
Sao Paulo, Brasil
E-mail: leandro.augusto@poli.usp.br

Rangaraj M. Rangayyan
University of Calgary
Department of Electrical and Computer Engineering
Schulich School of Engineering
and
Department of Radiology
Calgary, Alberta T2N1N4 Canada

Abstract. This work proposes a new approach using a committee machine of artificial neural networks to classify masses found in mammograms as benign or malignant. Three shape factors, three edge-sharpness measures, and 14 texture measures are used for the classification of 20 regions of interest (ROIs) related to malignant tumors and 37 ROIs related to benign masses. A group of multilayer perceptrons (MLPs) is employed as a committee machine of neural network classifiers. The classification results are reached by combining the responses of the individual classifiers. Experiments involving changes in the learning algorithm of the committee machine are conducted. The classification accuracy is evaluated using the area A under the receiver operating characteristics (ROC) curve. The A result for the committee machine is compared with the A results obtained using MLPs and single-layer perceptrons (SLPs), as well as a linear discriminant analysis (LDA) classifier. Tests are carried out using the student’s t-distribution. The committee machine classifier outperforms the MLP, SLP, and LDA classifiers in the following cases: with the shape measure of spiculation index, the A values of the four methods are, in order, 0.93, 0.84, 0.75, and 0.76; and with the edge-sharpness measure of acuteness, the values are 0.79, 0.70, 0.69, and 0.74. Although the features with which improvement is obtained with the committee machines are not the same as those that provided the maximal value of A (A = 0.99 with some shape features, with or without the committee machine), they correspond to features that are not critically dependent on the accuracy of the boundaries of the masses, which is an important result. © 2008 SPIE and IS&T. [DOI: 10.1117/1.2892683]

1 Introduction

Breast cancer is the most common type of cancer among women in many parts of the world. In Brazil, for example, the largest rate of incidence is among women between the ages of 35 and 50 years. In the United States, a constant increase in the rate of breast cancer incidence has been observed, mainly in women between 50 and 69 years of age.

The best manner to manage breast cancer is through early detection, which can be achieved via mammography, a special type of radiography of the breast for the purpose of diagnosing breast cancer. Mammography can detect tumors before they become palpable. Radiologists trained in mammography can achieve high accuracy in the diagnosis of breast cancer. However, studies have shown that the rate of wrong diagnosis by radiologists is of the order of 15 to 30%; such errors can lead to increased use of additional investigative procedures as well as increased numbers of unnecessary biopsies.

To minimize error in the diagnosis of breast cancer, systems have been developed for computer-aided diagnosis (CAD) of breast cancer. CAD systems, which use image processing to extract quantitative features from mammograms and pattern classification techniques for diagnostic decision, are based on the incorporation of clinical expertise of the radiologist into computational intelligence techniques.

Image processing techniques may be applied directly to a mammogram, or to suspicious regions, or to regions of interest (ROIs) to detect and extract characteristics of lesions in the breast. A lesion could be due to calcifications or masses that appear in the breast. With the objective to detect and extract the characteristics of lesions, several techniques of image processing have been proposed.

The artificial neural network (ANN) is an example of computational intelligence techniques that has been used to classify tumors related to breast cancer. Several types of network architecture, such as the multilayer perceptron (MLP), the single-layer perceptron (SLP), and ra-
dial basis functions (RBFs) have been used for the classification of breast masses and tumors based on measures of shape, texture, and edge sharpness.

In this work, we propose the use of a committee machine to classify breast masses as benign or malignant. The features of masses used include measures of shape, edge sharpness, and texture. We incorporate the learning algorithm for committee machines proposed by Murphey, Chen, and Feldkamp, which has been shown to provide improved classification accuracy. The classification accuracy is evaluated using the area under the receiver operating characteristics (ROC) curve. The results obtained with the committee machine are compared with those obtained using the MLP, the SLP, and linear discriminant analysis (LDA).

In the following sections, we describe the database of mammograms and the features used (Sec. 2); explain the nature of the committee machine of neural networks and the neural architecture (expert) used for composing the committee machine (Sec. 3); present the results obtained in the experiments conducted and compare the results with those of other classifiers such as the MLP, the SLP, and LDA (Sec. 4); present an analysis of the results using the hypothesis test with student’s t-distribution (Sec. 5); and finally provide a discussion of the results as well as concluding remarks (Sec. 6).

2 Features of Breast Masses

A set of 57 ROIs of mammograms, digitized with a resolution of 50 μm with 12 bits per pixel, of which 20 ROIs are related to malignant tumors and 37 ROIs to benign masses, is used in this work. The mammograms were obtained from Screen Test: the Alberta Program for the Early Detection of Breast Cancer (Alberta Cancer Board, Edmonton, Alberta, Canada). All cases with masses from a total of 170 cases (1745 mammograms) provided by Screen Test are included in the present study. The 170 cases were selected to build an indexed atlas of mammograms for other previous studies. A total of 57 ROIs related to masses were identified and the corresponding contours drawn by an expert radiologist specialized in screening mammography. All diagnoses were proven by biopsy.

In this work, we investigate the possibility of defining whether a mass is benign or malignant through attributes (quantitative characteristics) such as shape, edge sharpness, and texture. To facilitate the computation of the attributes, the boundaries (contours) of the masses were drawn by an expert radiologist experienced in mammography. Most benign masses have contours that are well circumscribed [Fig. 1(a)] smooth, and round or oval, although some could be macrolobulated [Fig. 1(b)]. On the other hand, malignant tumors typically exhibit ill-defined or blurred margins [Fig. 1(c)] as well as ill-differentiated and rough or spiculated contours [Fig. 1(d)]. However, some benign masses can have lobulated or spiculated contours and blurred margins, and some malignant tumors can have contours that are round or oval, as well as well-defined margins (see Fig. 1).

To represent the features of breast masses as described before, several researchers have been studying image processing techniques to extract measures of shape, texture, and edge sharpness. Quantitative features as described before have provided significant results in the discrimination between malignant tumors and benign masses in pattern classification studies. In this work, we use the features proposed by Rangayyan et al. and Alto, Rangayyan, and Desautels, which are described next.

Three shape factors were proposed by Rangayyan, Mudigonda, and Desautels to represent the shape complexity of breast masses: normalized compactness (C), fractional concavity (Fcc), and spiculation index (SI).

The feature C is a measure of shape complexity that has the value of 0 for a circle, typical of benign masses [see Fig. 2(a)]. In the case of irregular contours, typical of malignant tumors, it increases up to the maximum value of 1 [see Fig. 2(b)]. The normalized measure of C is given by:

\[ C = 1 - \frac{4\pi A}{p^2}, \]  

where A is the area enclosed and p is the contour of the perimeter.

The feature Fcc is computed by taking the cumulative length of the concave segments and dividing it by the total length of the contour. Benign masses [Fig. 2(a)] have fewer, if any, concave segments than malignant tumors [Fig. 2(b)]; thus, benign masses would have lower Fcc values than malignant tumors.

The feature SI represents the degree of spicularity of a mass contour. For the evaluation of this measure, Rangayyan, Mudigonda, and Desautels proposed a method based on a polygonal model of the given contour and a combination of the segment lengths, base widths, and angles of possible spicules. Benign masses are usually smooth or macrolobulated [Figs. 1(a) and 1(b)], and thus have lower values of SI compared to malignant tumors, which are typically microlobulated or spiculated [Figs. 1(c) and 1(d)]. Let \( S_n \) and \( \theta_n \), \( n = 1, 2, \ldots, N \), be the length and angle of N sets of polygonal model segments corresponding to the N spicule candidates of a mass contour. Then, SI is computed as:

\[ SI = \frac{\sum_{n=1}^{N} (1 + \cos \theta_n) S_n \sum_{n=1}^{N} S_n}{\sum_{n=1}^{N} \theta_n}. \]
The factor \((1 + \cos \theta_d)\) modulates the length of each segment (possible spicule) according to its narrowness. Spicules with narrow angles between 0 and 30 deg get high weighting, as compared to macrolobulations that usually form obtuse angles, and hence get low weighting.

The features \([C, F_{ic}, SI]\) are called “shape factors” in the following discussions.

Another set of features in this work is composed of edge-sharpness measures calculated across margins, defined as acutance \((A)\), contrast \((\text{contrast})\), and coefficient of variation \((CV)\).

The feature \(A\) is a measure of the sharpness or change in density (gradient) across a mass margin. It is computed using a line of pixels in the direction normal to the contour at each boundary pixel of a mass [see Figs. 2(c) and 2(d)]. A distance of 80 pixels or 4 mm inside and outside the mass contour is used where possible; in the case of narrow spicules or concavities, as many pixels as available are used, such that the line of pixels along the normal to the contour does not cross the contour more than once. The average of the differences between equidistant pixels on the inside and outside of the mass contour along the normal to each boundary pixel is computed, and further averaged over the entire contour to obtain a root-mean squared \((\text{rms})\) measure of edge sharpness. In addition, a measure of \textit{contrast}, as well as the \(CV\) of the average gradient along the normal to the contour are computed.

Benign masses tend to demonstrate sharper edges, and therefore, result in larger differences in pixel values between the pixels inside and outside a mass ROI [Fig. 2(c)], whereas malignant tumors tend to have more diffuse or ill-defined margins that result in smaller differences between pixel values [Fig. 2(d)]. Partially obscured mass margins or poorly drawn contours may result in lower values of \(A\). The \(A\) value could be expected to be high if a mass is mostly round, clearly visible, and denser than the surrounding breast tissue [Fig. 1(a)]. A highly spiculated mass, such as the example in Fig. 1(d), will generally have a poorly differentiated margin, where it is likely that the differences in pixel values across the boundary will be small; this will result in a low \(A\) value.

The features \([A, \text{contrast}, CV]\) are called “edge-sharpness measures” in the following discussions.

The third set of features used in this work is composed of texture measures. Based on the definitions of Haralick, Shanmugam, and Dinstein, \(28\) \(14\) texture measures \((f_j)\): angular second moment, \(f_2\); contrast, \(f_3\); correlation, \(f_4\); sum of squares, \(f_5\); inverse difference moment, \(f_6\); sum average, \(f_7\); sum variance, \(f_8\); sum entropy, \(f_9\); difference variance, \(f_{10}\); difference entropy, \(f_{11}\); information measures of correlation, \(f_{12}\); maximal correlation coefficient) were computed, using the pixels in adaptive ribbons of pixels surrounding the masses [Figs. 2(e) and 2(f)]. Mudigonda, Rangayyan, and Desautels \(14\) showed that texture features computed using mass margins ribbons surrounding the mass, [see Figs. 2(e) and 2(f)] could lead to improved discrimination between benign masses and malignant tumors as compared to texture measures computed using the entire mass region. The texture features were obtained after smoothing of the original images using a \(7 \times 7\) Gaussian kernel of standard deviation \(2\) pixels, and reduction from \(12\) bits to \(8\) bits per pixel. \(24\) The normalization of the gray-level co-occurrence matrix used in the computation of texture feature \(f_{14}\) was different from that given by Haralick, Shanmugam, and Dinstein, \(28\) as discussed in Ref. 29. In the following discussions, the full set of \(14\) texture features is called “texture measures.” Alto, Rangayyan, and Desautels \(24\) showed that the sum entropy measure (named \(f_1\)) is the most effective of the \(14\) texture features studied.

3 Committee Machine

The basic idea of the committee machine is to train a committee of ANNs to resolve a difficult computational intelligence task. In a committee machine, the individual results of each ANN (or expert) are combined to accomplish a better result of generalization. This process is related to the “divide and conquer principle.” \(30\) The computational simplicity is made possible by distributing a learning task among a number of experts (ANNs). The combination of the results provided by the experts composes a committee machine. One way to implement the committee machine is the AdaBoost algorithm, as proposed by Schapire \(31\) (see also Haykin \(30\)); this method is described next.
3.1 AdaBoost Technique

The AdaBoost (adaptive boost) technique gradually builds a composition of experts (ANNs), forming a committee with sequential learning.30,31 The aim of AdaBoost is to improve the accomplishment of a learning algorithm through an iterative process. In each iteration of the process, a new expert (an ANN) is added to the committee. We name the iteration index en, where en stands for the number of experts. The iteration is a progressive process starting with en = 1, where in each iteration a new expert is trained and added to the committee machine, up to a maximum of EN.

Given an ANN architecture used to implement the experts of the committee, which are named here NN1, NN2, …, NNen, …, NNEN, and a set of training patterns \( Tr = \{ (x_1, y_1), (x_2, y_2), ..., (x_N, y_N) \} \), where \( x_i \) belongs to some input domain \( X \), each label \( y_i \) belongs to the set \( Y \) of the output classes, and \( N \) is the number of training patterns, the AdaBoost algorithm presents a final hypothesis based on a combination of the hypothesis of each expert (ANN).

One of the main ideas of the AdaBoost algorithm is based on defining an adaptive set of probabilistic weights associated with the elements of the set of training patterns. The particular probabilistic weights applied to the set of training patterns in a given iteration \( en \) are called \( D_{en}(i) \), \( i = 1, 2, ..., N \). Initially, for \( en = 1 \), all probabilistic weights are set equal to the same value. Then, the weights are updated at each iteration \( en \) on the basis of the success or failure of the hypothesis \( h_{en-1} \) stated by the expert \( NN_{en-1} \).

When the expert \( NN_{en} \) receives a training set \( Tr_{en} \) with the probabilistic weights \( D_{en} \), it produces a hypothesis \( h_{en} : X \rightarrow \{ -1, 1 \} \). A sample of the training set that is misclassified by \( h_{en} \) will have its probabilistic weight increased when it is considered during the training of a new expert. In this manner, the AdaBoost algorithm increases the probabilistic weights of misclassified samples and reduces the probabilistic weights of correctly classified samples. In the next iteration, the expert \( NN_{en+1} \) is thus forced to focus on the difficult samples of the training set. This adaptive behavior is represented in Fig. 3.

The accuracy of the hypothesis of a given expert \( en \) is calculated by an error measure, as follows:

\[
\epsilon_{en} = \sum_{i=1}^{N} D_{en}(i)|h_{en}(x_i) - y_i|.
\]

Once the hypothesis \( h_{en} \) is produced by the expert, it is necessary to calculate a parameter \( \alpha_{en} \), given next, to measure the importance that is assigned to \( h_{en} \) in a linear combination of all the hypotheses of the several experts:

\[
\alpha_{en} = \frac{1}{2} \ln \left( \frac{1 - \epsilon_{en}}{\epsilon_{en}} \right).
\]

The process described before is continued until \( \epsilon_{en} \leq error_{max} \) or while \( en \leq EN \). At the end of this process, the committee machine is composed of the ANNs \( NN_1, NN_2, ..., NN_{en}, ..., NN_{EN} \) (see Fig. 4). The final hypothesis \( H(x) \) (produced by the committee machine) is given by

\[
H(x) = \text{sign} \left( \sum_{en=1}^{EN} \alpha_{en}h_{en}(x) - \text{CutPoint} \right),
\]

where \( \text{CutPoint} \) is a constant used to build the ROC curve.32

The architecture of the ANNs used in the committee machine is described next.

3.2 Artificial Neural Network or Expert

The ANN (or expert) used in the committee machine is the well-known MLP. The MLP experts are composed of three
layers of neurons (input layer, hidden layer, and output layer), with connections between the neurons in the different layers in the forward direction. An MLP with only one output neuron and one hidden layer, when used in a classification task given by a training set 

\[
\begin{align*}
\text{Tr} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_N, y_N)\},
\end{align*}
\]

where \(x_i \in \mathbb{R}^n\) and \(y_i \in \{-1, 1\}\), computes the function

\[
f_{\text{MLP}}(X) = \varphi \left[ \sum_{i=1}^{N_{hl}} v_i \cdot \varphi \left( \sum_{j=1}^{N_{hl}} w_{ij} x_j + b_{i0} \right) + b_{00} \right],
\]

where \(N_{hl}\) is the number of neurons in the hidden layer and \(N_{hl}\) is the number of neurons in the input layer. The synaptic weights are represented by \(v_i\) and \(w_{ij}\); \(b_{i0}\) and \(b_{00}\) are the biases; and \(\varphi(\cdot)\) is the activation function, which is commonly specified to be a sigmoid function.30

In an expert training process, we have to choose appropriate values for the parameters in the earlier expression to achieve minimal error, which can be calculated as the sum of the squared errors:

\[
E = \sum_{i=1}^{N} \frac{1}{2} [f_{\text{MLP}}(x_i) - y_i]^2.
\]

### 4 Experiments

To test the committee machine classifier, experiments were conducted with various combinations of features of 57 ROIs of breast masses. The features described in Sec. 2 were combined in various sets, as listed in Table 1, to compare the committee machine, as proposed in the present work, with other classifiers such as the MLP, the SLP, and LDA.19,24 The results of performance of the classification experiments are shown in terms of the area under the ROC curve \(A_z\).33 Illustration of one set of ROC curves is provided in Fig. 5.34 To obtain each \(A_z\) value, calculated as described by Metz33 and shown in Table 1, we conducted a series of 20 experiments. In each of these experiments, the features of 20 randomly chosen ROIs (ten malignant and ten benign) were used to form the test set, and the remaining ROIs (ten malignant and 27 benign) were used as the training set. The test and training sets were mutually exclusive in each experiment. Each result shown in Table 1 provides the mean and the standard deviation of the \(A_z\) values obtained in the 20 experiments.

In our first study on this subject,22 we presented initial results obtained with the committee machine, MLP, and SLP, using only the shape and edge-sharpness features listed before. In that study, a new learning process was initiated at each iteration in each new expert (ANN). For this reason, the learning time was long. To minimize...
the learning time without losing accuracy, in our second study,\textsuperscript{23} we implemented a continuous learning process as proposed by Murphey, Chen, and Feldkamp,\textsuperscript{25} where the knowledge acquired by an expert is used by the next one. This is accomplished by using the synaptic weights obtained by one expert (final $w_{ij}$ and $v_i$ after learning) as the initial knowledge (initial $w_{ij}$ and $v_i$) of the next expert. We conducted two series of experiments with the committee machine: one with the traditional method, in which the learning process is initiated with random synaptic weights for each new iteration (series named “without continuous W,” where the vector W represents the synaptic weights), as in Silva, Del-Moral-Hernandez, and Rangayyan,\textsuperscript{26} and the other series of experiments utilizing the knowledge of the previous expert (“with continuous W”). The use of the continuous learning process improved the results in some cases, but also led to slight reductions of the $A_e$ values in other cases.

In most cases of feature combinations, the committee machine outperformed the MLP, SLP, and LDA methods. Of particular interest are the improved results provided by the committee machine with edge-sharpness measures: these features had previously been observed by Alto, Rangayyan, and Desautels\textsuperscript{24} to result in low classification accuracy as compared with the shape features using other classification procedures (LDA, logistic regression, and the Mahalanobis distance). The results obtained in this work demonstrate that the committee machine can provide good results with weak features that may not yield good results with other classification methods.

For the single MLP classifier topology, the following configuration was adopted: the number of neurons in the input layer ($N_{il}$) is the same as the number of features used, for example, shape factors ($N_{il}=3$), edge-sharpness measures ($N_{il}=3$), and texture measures ($N_{il}=14$). To select the number of neurons in the hidden layer ($N_{ih}$), we conducted experiments to verify the learning curve (the mean-squared error or $MSE$) against the number of iterations) as explained by Haykin,\textsuperscript{30} using $N_{il}$=5, 10, 15, 20, and 30. Based on the results of these experiments, we decided to use $N_{il}=30$ because it resulted in a low $MSE$ with a small number of iterations in all experiments with the different feature sets. In the output layer, we used only one neuron ($N_{ol}=1$) to represent the malignant or benign class. Thus, the MLP topology in the prior examples is $N_{il}=3$, 3, or 14 (depending on the feature set used), $N_{il}=30$, and $N_{ol}=1$. The maximum $MSE$ allowed was 0.2, and the maximum number of iterations allowed was 10,000.

As described in Sec. 3, the experts of the committee machine classifier were implemented through MLPs. The topology of the MLP experts is less complex than the topology in the single MLP classifier. This can be done because, under certain limits, the reduction of complexity of the experts does not reduce significantly the global performance of the committee. In particular, the number of nodes in the hidden layer of the MLP experts was fixed at ten neurons ($N_{ih}$=10). The maximum $MSE$ error allowed for the MLP experts was fixed at $error_{max}=0.35$. Table 2 provides a description of the number of experts (MLP) used in the training process of the committee machine.

Note that even for committee machines with large numbers of experts and large numbers of features, the maximal time required for training is less than 200 s. This is the case for a conventional and inexpensive computer: Athlon XP 2.2 GHz with 512 MB of RAM. This range of training time is acceptable for our type of application.

For the SLP, the training procedure was conducted until $error=0.2$, and the maximum number of iterations allowed was 10,000.
5 Statistical Analysis of the Results

Table 1 shows the mean and the standard deviation of the $A_c$ values obtained in 20 experiments for each set of features listed. To derive statistical measures of comparison of the results obtained in the experiments, we conducted hypothesis tests using the student’s $t$-distribution, as described by Spiegel.\(^{35}\) In the present study, the number of experiments conducted to obtain each $A_c$ is 20; therefore, the number of degrees of freedom considered is 38.

The value of $t$ as a function of the number of degrees of freedom and the desired confidence interval can be found in Spiegel.\(^{35}\) Based on a bilateral test, with the confidence interval of 0.05, the null hypothesis is rejected when $t$ is outside the interval $-t_{0.025}$ to $t_{0.025}$, for which, with 38 deg of freedom, the tabulated values are $t_{tab} = -2.02$ to $t_{tab} = 2.02$. Consequently, for any value outside this interval, the results of the classifiers being compared are statistically different.

Some of the results of the hypothesis test using the student’s $t$-distribution are shown in Fig. 6 for illustration. A complete statistical analysis of the performance of the classifiers tested is shown in Table 3. The results in Table 3 relate to the committee machine without continuous $W$ in a comparative study with other classifiers.

Notice that the contrast based on the hypothesis test described before, which shows the advantage of the committee machine for the classification based on the features $[SI, A]$, $[SI]$, and $[A]$, focuses on the contrast between the averages of $A_c$. To have better contrastive results, more detailed experiments were carried out for these three sets of features, and new contrast calculations, targeting not only the averages of $A_c$ but also taking into account the dispersions of $A_c$, were done. Specifically, for each random choice of a training set, we had the measure of the difference $\Delta A_c$, between the $A_c$ of the committee machine and the $A_c$ of the MLP classifier (the more powerful choice after the committee machine). Based on the average and the standard deviation of this variable $\Delta A_c$, we can estimate the probability of a committee machine having a larger $A_c$ than the MLP classifier, since this corresponds to the probability of $\Delta A_c$ being positive. This leads us to the experimental results summarized in Table 4, which indicates, for example, that the use of combined features $[SI, A]$ results in 88% of the random training sets leading to better $A_c$ values for the committee machine.

### 6 Discussion and Conclusion

The CAD systems discussed in this work use two complementary techniques, image processing and computational intelligence. Techniques of image processing are used in this work to extract features to represent the differences between malignant tumors and benign masses. The extracted features are used in a committee of ANNs to perform classification. The classification results are analyzed using the area under the ROC curve and compared with results obtained using other classifiers.
It should be noted that, while the shape features used provide high classification accuracy, they depend on the availability of accurate boundaries of masses, which are hard to draw even by an expert radiologist, and harder to obtain by the application of image processing algorithms. Preliminary results obtained using automatically detected contours of masses according to the methods proposed by Mudigonda, Rangayyan, and Desautels\(^\text{15}\) indicated that the contours were not adequately accurate to permit the derivation of shape features with good classification performance. However, Sahiner et al.\(^\text{11}\) proposed automatic methods to obtain contours of masses, with which the highest classification accuracy was provided by a shape factor based on Fourier descriptors, as compared to several morphological and texture features. Further evaluation of the performance of the various features used in the present work is desirable, using contours of masses drawn by several radiologists and obtained automatically by many image processing algorithms.

The edge-sharpness and texture features used in the present work are derived using ribbons of pixels around the mass ROIs; for this reason, they are not critically dependent on the accuracy of the boundaries of the masses. However, the classification performance of the edge-sharpness and texture features, using traditional pattern classification methods such as LDA and logistic regression, has been found to be poor.\(^\text{24}\) Features that perform poorly in pattern classification experiments using well-established, linear pattern classification procedures, such as edge sharpness and texture, need not be rejected if they are deemed important from other perspectives, such as radiological, pathological, and physiological points of view, i.e., if they carry important discriminative information from clinical perspectives.\(^\text{36,37}\) If that is the case, then the poor results can be related to limitations of the classifier, and not necessarily to limitations of the discriminative power of the features. For this reason, the classification performance of such features could be improved via the use of advanced classifiers, such as committee machines composed of several ANNs. The combined use of features representing several different radiological characteristics, such as edge sharpness, texture, and shape, instead of the use of only shape features, could lead to improved representation and analysis of breast masses in mammograms.

To verify the efficiency of the committee machine in the classification of malignant tumors and of benign masses, comparative performance experiments using others classifiers (MLP, SLP, and LDA) were conducted. Statistical analysis of the classification results was performed to verify if the committee machine outperformed the other classifiers (summary in Table 3). When we compare the experimental results of the MLP and the committee machine, we see that we have performance improvements in some cases of grouping of features (Fig. 6). Our main goal of the present work and application, i.e., improvement of classification...
accuracy, is achieved in such cases. Note that one of these cases corresponds to A, an edge-sharpness feature, which is less dependent on the accuracy of the boundaries of masses than shape factors, as mentioned before. Note also that, although the search for higher classification rates through committee machines results in large training and processing times, these times are limited to reasonable amounts (less than 200 s in the worst case). From the experiments conducted and results obtained, we can also verify the attributes that can efficiently discriminate between a malignant tumor and a benign mass (summary in Table 1). When using different classifiers with different capabilities of discrimination, through the calculation of A, it is possible to select the characteristics or sets of features that perform better than others. The results of our study should be useful in the design and evaluation of CAD systems.

Acknowledgment
The authors thank the University of São Paulo (Brazil), the Polytechnic School of the University of São Paulo (Brazil), and the National Council for Scientific and Technological Development (CNPq, Brazil) for supporting this work. They also thank the Catalyst Program of Research Services, University of Calgary, for supporting this collaborative project.

References
34. ROCKIT, see http://www.radionomics.uchicago.edu/kt/krhoc/software/index.htm.

Leandro Augusto da Silva is an electrical engineer from the School of Industrial Engineering (FEI), with a MSC in electronics systems from the Polytechnic School from the University of Sao Paulo, Brazil. He is currently a PhD student in the Department of Electronic Systems Engineering of the Laboratory of Integrated Systems (LSI) at the Polytechnic School from the University of Sao Paulo, Brazil and researcher of a Group Division of Informatics at the Heart Institute at Medical School at University of Sao Paulo (InCor).

Journal of Electronic Imaging 013017-9 Jan–Mar 2008/Vol. 17(1)
Emilio Del Moral Hernandez received BSc and MSc degrees in electrical engineering from the Polytechnic School of the University of Sao Paulo, Brazil, and the PhD degree in electrical engineering from the University of Pennsylvania, Philadelphia. He is currently with the Department of Electronic Systems Engineering, University of Sao Paulo, where he is an associate professor. His research activities include the fields of computational intelligence, neuro-like computation and biologically inspired information processing, complex systems and nonlinear dynamics, modeling of nonlinear phenomena, pattern recognition, classification and clustering, data mining, digital signal processing, optimization, automatic decision, the development of novel neurocomputing architectures that explore bifurcation and chaos, pulsed model neurons and electronic implementation, hybrid architectures blending more than one classical neural architecture, and data fusion and multisensors systems for environmental and medical applications.

Rangaraj M. Rangayyan is a professor with the Department of Electrical and Computer Engineering, and an adjunct professor of surgery and radiology at the University of Calgary, Alberta, Canada. He received the bachelor of engineering degree in electronics and communication in 1976 from the University of Mysore at the People’s Education Society College of Engineering, Mandya, Karnataka, India, and the PhD degree in electrical engineering from the Indian Institute of Science, Bangalore, Karnataka, India, in 1980. His research interests are in the areas of digital signal and image processing, biomedical signal analysis, and biomedical image analysis. He is the author of two books: Biomedical Signal Analysis (IEEE/Wiley, 2002) and Biomedical Image Analysis (CRC, 2005). He was recognized by the IEEE with the award of the Third Millennium Medal in 2000, and was elected as a Fellow of the IEEE in 2001, Fellow of the Engineering Institute of Canada in 2002, Fellow of the American Institute for Medical and Biological Engineering in 2003, and Fellow of SPIE in 2003.