# A Multiple Circular Path Convolution Neural Network System for Detection of Mammographic Masses

Shih-Chung B. Lo*, *Member, IEEE*, Huai Li, *Member, IEEE*, Yue Wang, *Member, IEEE*, Lisa Kinnard, and Matthew T. Freedman

***Abstract*—A multiple circular path convolution neural network (MCPCNN) architecture specifically designed for the analysis of tumor and tumor-like structures has been constructed. We first divided each suspected tumor area into sectors and computed the defined mass features for each sector independently. These sector features were used on the input layer and were coordinated by convolution kernels of different sizes that propagated signals to the second layer in the neural network system. The convolution kernels were trained, as required, by presenting the training cases to the neural network.**

**In this study, randomly selected mammograms were processed by a dual morphological enhancement technique. Radiodense areas were isolated and were delineated using a region growing algorithm. The boundary of each region of interest was then divided into 36 sectors using 36 equi-angular dividers radiated from the center of the region. A total of 144 Breast Imaging—Reporting and Data System-based features (i.e., four features per sector for 36 sectors) were computed as input values for the evaluation of this newly invented neural network system. The overall performance was 0.78–0.80 for the areas ($A_z$) under the receiver operating characteristic curves using the conventional feed-forward neural network in the detection of mammographic masses. The performance was markedly improved with $A_z$ values ranging from 0.84 to 0.89 using the MCPCNN. This paper does not intend to claim the best mass detection system. Instead it reports a potentially better neural network structure for analyzing a set of the mass features defined by an investigator.**

Manuscript received February 22, 2000; revised January 11, 2002. This work was supported by the US Army under Grant DAMD17-96-1-6254 through a subcontract from University of Michigan, Ann Arbor, and under Grant DAMD17-01-1-0267 through a subcontract from Howard University. The work of Y. Wang was supported by the US Army under Grant DAMD17-98-1-8045. The work of L. Kinnard was supported by the US Army under Grant DAMD 17-00-1-0291. The content of this paper does not necessarily reflect the position or policy of the government. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was N. Karssemeijer. *Asterisk indicates corresponding author*.

*S.-C. B. Lo is with the Center for Imaging Science and Information System, Radiology Department, Georgetown University Medical Center, 2115 Wisconsin Avenue, Suite 603, N.W., Washington, DC 20007 USA (e-mail: lo@isis.imac.georgetown.edu).

H. Li was with the ISIS Center, Radiology Department, Georgetown University Medical Center, Washington, DC 20007 USA. He is now with the Center for Information Technology, Division of Computational Bioscience, National Institutes of Health, Bethesda, MD 20892 USA.

Y. Wang is with the Department of Electrical Engineering and Computer Sciences, The Catholic University of America, Washington, DC 20064 USA.

L. Kinnard is with the Center for Imaging Science and Information System, Radiology Department, Georgetown University Medical Center, Washington, DC 20007 USA, and also with the Department of Electrical Engineering, Howard University, Washington, DC 20059 USA.

M. T. Freedman is with the Center for Imaging Science and Information System, Radiology Department, Georgetown University Medical Center, Washington, DC 20007 USA.

Publisher Item Identifier S 0278-0062(02)02935-X.

***Index Terms*—BI—RAD, computer-aided diagnosis, convolution neural network, mammography masses, neural network, sector features.**

## I. Introduction

IT IS KNOWN that effective treatment of breast cancer calls for early detection of cancerous lesions (e.g., clustered microcalcifications and masses associated with malignant cellular processes) [1]–[3]. Breast masses appear as areas of increased density on mammograms. It is particularly difficult for radiologists to detect and analyze a suspected area where a mass is overlapped with dense breast tissue. These masses are more readily seen as time progresses, but the further the tumor has progressed, the lower the possibility of a successful treatment. Therefore, increasing the chances of early breast cancer detection in improving today's clinical system is of vital importance in breast cancer diagnosis.

Several research groups have developed computer algorithms for automated detection of mammographic masses [4]–[8]. Some of these methods involved in classification of masses and normal dense breast tissues [7], [8]. Investigators also attempted to classify the malignant or benign nature of the detected tumors [9]–[11]. It is conceivable that correct segmentation of the masses [12] plays an important processing step prior to further mass analysis. In short, the results of these detection programs indicate that a high true-positive (TP) rate can be obtained at the expense of two or three false-positive (FP) detections per mammogram. Mammographically, a multiplicity (more than two) of similar benign-appearing breast lesions argues strongly for benignity [13]–[16] and, indeed, the more masses that are identified, the less chance that they represent cancer [17]. If the computer indicates multiple suspicious locations on a mammogram, the radiologist has to seek out one mass that possesses mammographic features, which are different from the others. The significant lesion may be missed due to the multiplicity of possible lesions. We, therefore, believe that a more useful and fundamental approach to computer-aided diagnosis (CAD) of masses is to devise computer programs to analyze features of a suspected area [18], [19] and to provide feature measures and estimates of the likelihood of malignancy by making comparisons within a digital mammographic database. The computer, therefore, serves as a second opinion and also provides a reproducible and an objective evaluation of the mass. With this aid, the radiologist may also increase his/her sensitivity by lowering the threshold of suspicion, while maintaining the overall specificity and reading efficiency.

## II. Clinical Background of Breast Lesions and Technical Approach in Mass Detection

### A. Description of Clinical Background

Most commonly, breast cancer presents itself as a mass. The same lesion shows a somewhat different picture from one projection to the other. Difficulties in masses also vary with the underlying breast parenchyma. In the fatty breast, masses are generally easy to detect. In the dense breast, mass detection is more difficult and auxiliary signs aid this detection. When the breast contains one mass, the decision process is based on its size, shape, and margins. When there are several masses, one looks at each, trying to determine whether any has features to suggest cancer. Furthermore, one looks to see if any mass is different in appearance from the others. Multiple small, well-defined, similar masses that present themselves bilaterally are all likely to be benign. Large, poorly defined, spiculated and unusually radiodense masses are extremely likely to be malignant. In this study, we used several computational features (see Section III-B) highly associated with four major features of breast masses routinely used in clinical reading:

*Density:* Malignant lesions tend to have greater radiographic density due to high attenuation and less compressibility of cancer than normal tissue. Radiolucent lesions are typically benign and the diagnosis can be made from the mammogram.

*Size:* If the lesion has morphological features suggesting malignancy, it should be considered suspicious regardless of the size. Isolated masses with noncystic densities greater than 8 mm in diameter can be malignant. In general, the larger a lesion, the more suspicious it is.

*Shape:* The more irregular the shape of a lesion, the more likely the possibility of malignancy. Lesions tend to be round, ovoid and/or lobulated. Small and frequent lobulations are suspicious. Lesions in the lateral aspect of the breast near the edge of the parenchyma with a reniform shape and a hilar indentation or notch usually represent a benign intramammary lymph node. Breast carcinoma hidden in the dense tissues can cause parenchymal retraction, which possess different shapes.

*Margins:* The margins of the lesion should be carefully evaluated for areas of spiculation, stellate patterns or ill-defined regions. Most breast cancers have ill-defined margins secondary to tumor infiltration and associated fibrosis. The appearance of spiculations and a more diffuse stellate pattern are almost pathognomonic for cancer. Lesions with sharply defined margins have a high likelihood of being benign; however, up to 7% of malignant lesions can be well circumscribed.

These are known clinical features and have been adapted in "Breast Imaging—Reporting and Data System" (BI—RAD) [20] of the American College of Radiology. Fig. 1(a) and (b) shows two breast images containing masses. In Fig. 1(a), a malignant mass is superimposed on the dense glandular tissue.

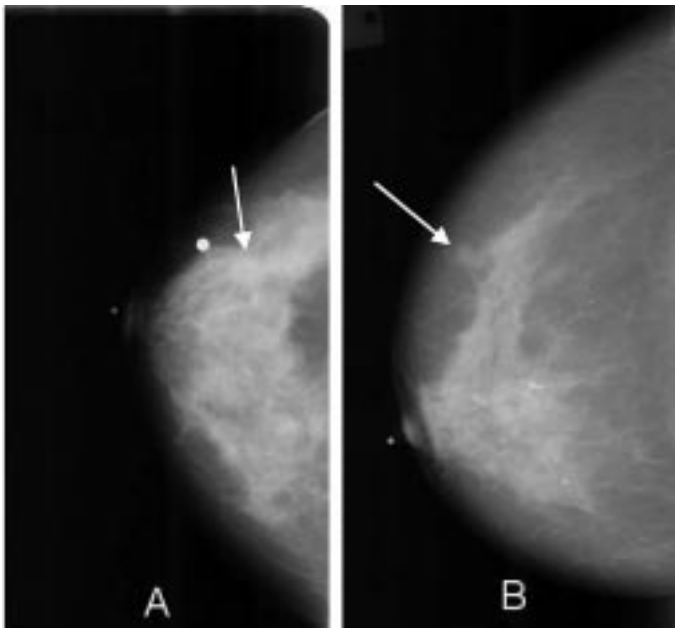


Fig. 1. (a) Dense breast containing a malignant mass. (b) Fatty and glandular breast containing a malignant mass.

However, its spiculated nature makes it easily identifiable. In Fig. 1(b), another malignant mass is located on the fatty background but is associated with a large body of glandular tissue. This mass is not easily detectable by the computer because its density is lower than the neighboring glandular tissue. Furthermore, one end of the mass is fully connected with this tissue.

### B. Technical Approach for Detection of Mammographic Masses

In this study, our goal was to detect clinically suspicious lesions. The differentiation of benign and malignant status of the mammographic masses can be extended from this study model and will be reported in our future work. The study was conducted with the following steps: 1) use background correction method and morphological operations to extract radio-opaque areas; 2) delineate the boundary of the areas; 3) compute the features and texture of the masses with emphasis on the boundary; and 4) design training strategy using neural networks as classifiers for the recognition of mass features. The overall detection scheme of the study framework is shown in Fig. 2.

## III. Development of Technical Methods

### A. Preprocessing and Extraction of Suspicious Masses

In automatic mass detection, accurate selection of suspected masses is considered a critical first step due to the variability of normal breast tissue and the lower contrast and ill-defined margins of masses. In our previous study [18], we aimed to improve the task of lesion site selection using model-based image processing techniques for unsupervised lesion site selection. We focused on two essential issues in the stochastic model-based image segmentation: enhancement and model selection. Based on the differential geometric characteristics of masses against

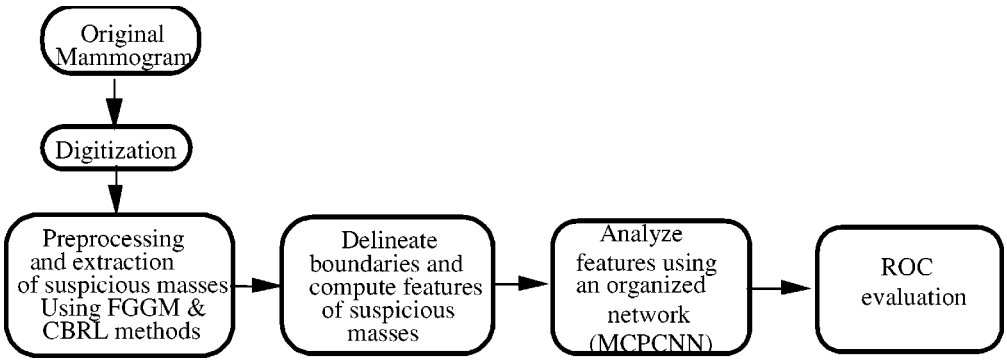


Fig. 2. A system flow chart for the detection of masses in this study.

the background tissues, we proposed one type of morphological operation to enhance the mass patterns on mammograms by removing high intensity background caused by breast tissues while maintaining mass-signals [18]. Then we employed a finite generalized Gaussian mixture (FGGM) distribution to model the histogram of the mammograms where the statistical properties of the pixel images are largely unknown and are to be incorporated. We incorporate the expectation-maximization algorithm with two information theoretic criteria to determine the optimal number of image regions and the kernel shape in the FGGM model. Finally, we applied a contextual Bayesian relaxation labeling (CBRL) technique to perform the selection of suspected masses.

We consistently processed the mammograms using this prescreening segmentation method. In the previous study [18], the FGGM method isolated 1142 potential masses including 114 of the 186 true masses in 200 mammograms. The mammograms were collected from the Mammographic Image Analysis Society (MIAS) database [21] and Brook Army Medical Center (BAMC) database. After morphological enhancement, 3143 potential masses were extracted using the FGGM technique. Of them, 181 were masses; however, five masses were not extracted. The results demonstrated that more true masses were picked up after enhancement although more false cases were also included. The undetected areas mainly occurred at the lower intensity side of the shaded objects or more obscured by fibroglandular tissues that, however, were extracted on morphological enhanced mammograms. Additionally, when the margins of masses are ill defined, only parts of suspicious masses were extracted from the original mammograms. We, therefore, decided to use the proposed morphological operation as a preprocessing step for the image enhancement prior to a segmentation method for the extraction of potential masses on the mammograms.

Based on the CBRL segmented region of interest (ROI), we employed a region growing method using a four-neighbors connection method assisted with a template masking operation to fill unconnected holes in the ROI

$$\text{IF } f(x-a, y-b) > V \text{ and } f(x, y) \in S, \text{ then } f(x-a, y-b) \in S \quad (1)$$

$$\text{IF } f(x-d, y-d) \in S, \text{ then } f(x-t, y-s) \in S \text{ for } t \le d \text{ and } s \le d \quad (2)$$

where $V$ denotes the threshold value of the originally CBRL segmented ROI, $S$ represents the set of growing region, and $[a, b]$ is a set of four conditions (i.e., [1, 0], [−1, 0], [0, 1], and [0, −1]) for the four neighboring pixels. In (2), $d$ is the size of template. In practice, we found that $d$ should be set at five pixels to fill the holes without disrupting the boundary.

### *B. Feature Extraction of the Masses*

Feature extraction methods play an essential role in many pattern recognition tasks. Once the features associated with an image pattern are extracted accurately, they can be used to distinguish one class of patterns from the others. Recently, many investigators have found that the multilayer perceptron (MLP) neural network using the error backpropagation training technique is a very powerful tool to serve as a classifier [22], [23]. In fact, the use of MLP neural network system for classification of disease patterns has been widely applied in the field of CAD [24]–[28].

The success of using a classifier for a pattern recognition task would rely on two factors: 1) selected features that could describe a discrepancy between image patterns and 2) accuracy of the feature computation. Should either one fail, no analyzer or classifier would be able to achieve an expected performance. By analyzing many clinical samples of various sizes of masses, we found that the peripheral portion of the mass plays an important role for mammographers to make a diagnosis. The mammographer usually evaluates the surrounding background of a radiodense area when a region is suspected.

We used the CBRL segmented ROI to compute the center. Since the segmented ROIs were somewhat smaller than the mammographer's delineation and on the denser region of the suspected patch, the computed centers were quite close to the visual center. We then divided the boundary of the ROI into 36 sectors (i.e., $10^o$ per sector) using 36 equi-angular dividers radiated from the center of the ROI. The following features were computed within each $10^o$ sector of the region.

a) "$l$"—the length from the center of the ROI to the boundary segment of the sector.
b) "$a$"—the cos($\theta$) (where $\theta$ is the normal angle of the boundary).
c) "$g$"—the average gradient of gray value on the segment along the radial direction (i.e., $g = \sum_{i=1}^{N}\{g_i/N\}$) where $N$ is the number of pixels of $i$ along the radial direction from $l/3$ inside the boundary to the boundary (see the left

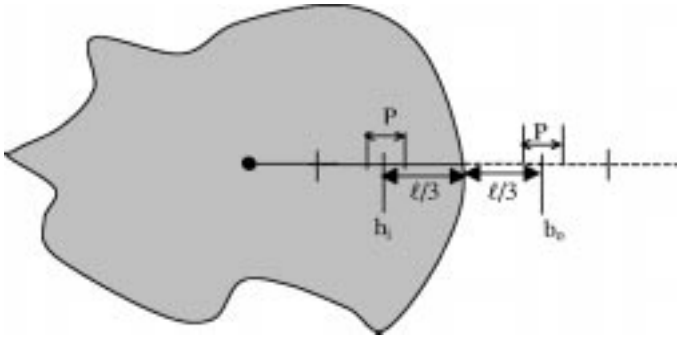


Fig. 3. A suspicious mass is delineated and shown as the shaded region. Contrast is computed by subtracting the average background pixel value (i.e., $b_o$, $o = 1, 2, \ldots P$) from the average foreground value (i.e., $h_i$, $i = 1, 2, \ldots P$).

$l/3$ line segment, Fig. 3). Technically speaking, this set of gradient values may also serve as a fuzzy system on the input layer in the neural network (to be described in Section III-C).

d) "$c$"—the gray value difference (i.e., contrast) along the radial direction. Specifically, $c = \sum_{i=1}^{P}\{h_i/P\} - \sum_{o=1}^{P}\{b_o/P\}$ where $h_i$ (or $b_o$) represents a pixel value along the radial direction. The position $l/3$ inside the boundary is the center of pixels $h_i$ ($i = 1, 2, 3, \ldots P$) and position $l/3$ outside the boundary is the center of pixels $b_o$ ($o = 1, 2, 3, \ldots P$), and $P$ is the number of pixels equivalent to a segment of $l/6$ and was used for averaging (see Fig. 3).

Hence, a total of 144 computed features (four features/sector for 36 sectors) were used as input values for the classification of the ROI. The relationship between the computed features and BI—RADS descriptors are discussed below.

i) ROI Size—The size of ROI is provided by the 36 "$l$" values.
ii) ROI Shape (round, oval, lobulated, or irregular)—The 36 "$l$" and 36 "$a$" values can describe the shape of the ROI.
iii) ROI Margin (circumscribed, microlobulated, obscured, ill- defined, or spiculate)—The 36 "$g$" and 36 "$l$" values can describe the ROI margin.
iv) ROI Density (fat-containing, low density, isodense, or highly dense)—The 36 "$c$" and 36 "$g$" values can be used to describe the density distribution of the ROI.

In short, the selected features are greatly associated with the main mass descriptors indicated in the BI—RADS. The reason for using 36 values for each nominated feature is four-fold: 1) mass boundary varies, it is difficult to describe an image pattern using a single value; 2) due to the general shape of the masses, the features of masses can be easily analyzed by the polar coordinate system; 3) in case some features are inaccurately computed in several directions due to the structure noises, such as the breast slender lines, there may still exist a sufficient number of correct features; and 4) generally more accurate results can be produced by using subdivided parameters rather than using global parameters in a pattern recognition task when the parameters are barely discernable and sample sizes are sufficiently large. Other computational features (e.g., difference entropy [19] and other higher order features) are eligible but require further investigation.

### *C. The Neural Network Structure Specifically Designed for the Extracted Boundary Features*

*1) Multiple Paths With Circular Networking to Instruct the Neural Network in Analyzing Sector Features:* This paper focuses on neural network design and arrangement of features for effective pattern recognition of ROIs. We designed several neural network connections between the input and the first hidden layers as shown in Fig. 4. In this neural network system, the first layer also functions as a correlation layer that transforms and encodes the signals from input nodes into correlation features for further neural network process. Fig. 4(a)–(c) illustrates the full connection (FC), a self correlation (SC) network, and a neighborhood correlation (NC) network, respectively. Network connections with multiple sectors (i.e., $20^o$, $30^o$, $40^o$, and $50^o$ of the NC) are grouped separately as independent NC paths. In the following study, we used four SC paths for a single sector and thirteen NC paths for four types of multisectors. The method of using the multiple correlation connections was motivated by our research experience in two-dimensional (2-D) convolution neural network (CNN) [(2-D CNN)] where we found that more than ten multiple convolution kernels in the CNN were necessary in the detection of lung nodules and microcalcifications [25].

Compared with 2-D CNN systems, the computation required in the one-dimensional (1-D) CNN (e.g., 144 input features) is relatively small. The combination of the networking paths described earlier for multiple circular path convolution neural network (MCPCNN) was implemented using C programming language. The internal computation algorithm used in the MCPCNN shares the same convolution process as that in the 2-D CNN [25]. Rotation invariance and flip invariance for training the 1-D convolution kernels in the MCPCNN were employed.

The fully connected neural network is a conventional feedforward MLP neural network. The signals of the fully connected neural network join the other network processes (i.e., SC paths and NC paths) at the single node of the output layer. The signal received at the output node is scaled between zero and one. During the training, zero and one were assigned at the output node to perform backpropagation computation for a nonmass and a mass, respectively. The backpropagation is computed in such a way that the computed incremental errors [see equations (9) and (10)] are retraced into every independent network path. Excluding the output layer, the SC and NC signals are independently arranged and are processed through the 1-D convolution process in the forward propagation. The learning algorithms for all three types of circular network paths are based on the backpropagation training method.

Let $V^0(n', s')$ represents an input signal at the node $n'$ and sector $s'$. The signal processed through an NC path and to be received at each node, $n$, on the first hidden layer is

$$N^1_{j[NC]}(n) = \left[\sum_{s'}\sum_{n'} V^0(n', s') \cdot W_{j[NC]}(n', s'; n)\right] + b^0_{j[NC]}(n) \quad (3)$$

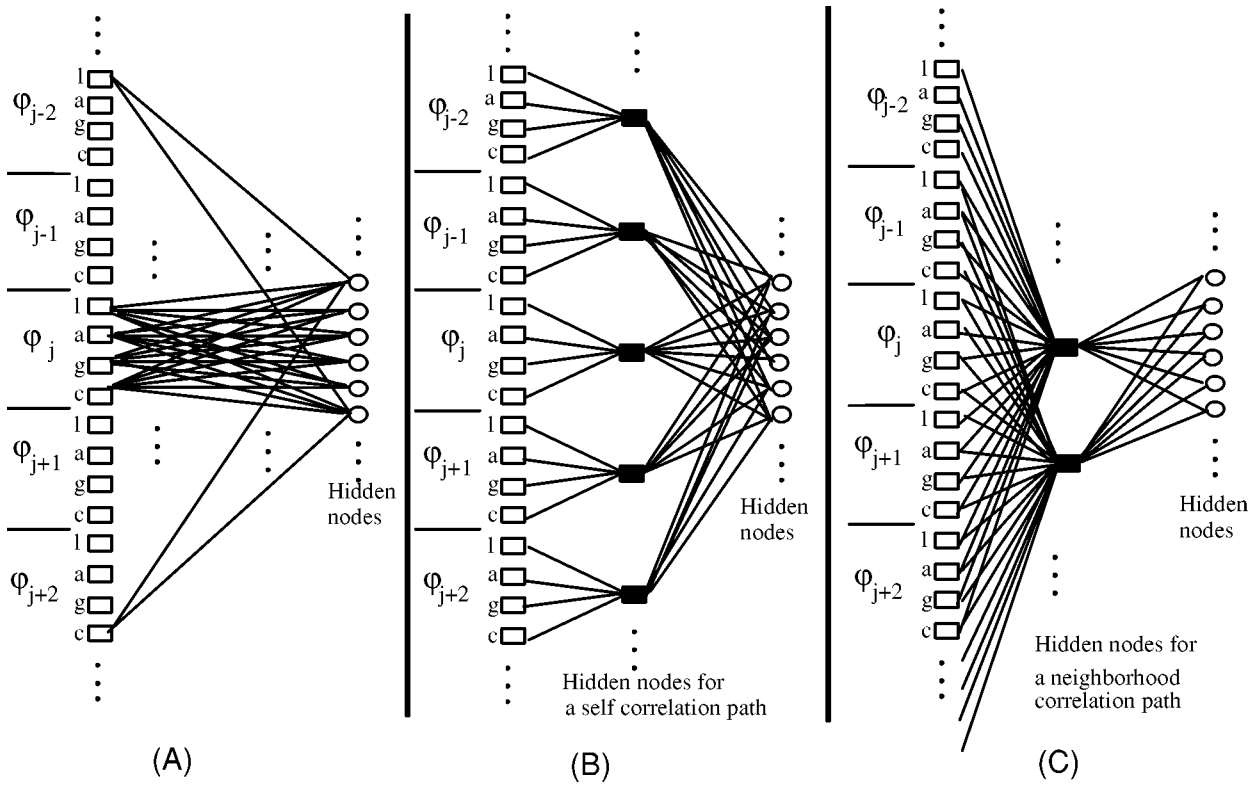


Fig. 4. Three types of network paths connecting the input and the hidden layers in the MCPCNN. (a) FC path. (b) SC path. Eeach node on the layer connects to a single set of the features (l, a, g, c) for the fan-in and fully connects to the hidden nodes for fan-out. (c) A NC path. Each node on the layer connects to the input nodes of adjacent sectors for the fan-in and fully connects to the hidden nodes for fan-out. The fan-in nets emphasizing SC in (b) and NC in (c) represent convolution weights (i.e., the same type of sectors possess the same set of weighting factors).

where $b^0_{j[NC]}(n)$ represents the bias term and $W_{j[NC]}(n', s'; n)$ is an array associated the 2-D nets that fan-in to a given receiving node, $n$. Each element of $W_{j[NC]}(n', s'; n)$ is the weight factor connected to node $n$ from node $n'$ sector $s'$ through a NC path, $j$, and $s'$ covers a range of neighborhood sectors corresponding to each type of NC path. Note that multiplications between the input nodes and connecting weights are computed first followed by taking the sum of the products for those nodes and sectors involved. The operation is repeated by shifting the weights from one set of sectors to the next. The procedure involving array multiplication passing through every sector is referred as the 1-D convolution operation that takes place in the sector dimension. The signal processed through an SC path and to be received at a node, $n$, on the first hidden layer is a special case of an NC path when $s'$ only covers one sector

$$N^1_{i[SC]}(n) = \left[\sum_{n'} V^0(n',\, s') \cdot W_{i[SC]}\,(n'; n)\right] + b^0_{i[SC]}(n) \tag{4}$$

where $W_{i[SC]}(n'; n)$ is the weight factor connected to $n$ from node $n'$ through a SC path, $i$, regardless of the sectors. A total of 18 paths (1 FC, 4 SC paths, and 13 NC paths for four types of multisectors) were used in our experiment described later. Nevertheless, the signals processed through a path and to be received at each node, $n$, on the first hidden layer is

$$V^1_P(n) = S\left(N^1_P(n)\right) \tag{5}$$

where $p$ is one of the network paths and $S(z)$ is a sigmoid function given by

$$S(z) = \frac{1}{1 + \exp\,(-z)}. \tag{6}$$

The sigmoid function would produce modulated values ranging from zero to one. The signals on other hidden layers in each path are processed the same as a conventional fully connected neural network. Other than the first hidden layer, the receiving signals at a hidden layer, $l$, collected from the previous hidden layer, $l$ to one, are merged from the nodes in the last layer and are given by

$$\begin{aligned} V^l(n) &= S\left(N^l(n)\right) \\ &= S\left(\sum_{n'} V^{l-1}(n') \cdot W^{l-1}(n'; n) + b^{l-1}(n)\right) \end{aligned} \tag{7}$$

where $n'$ and $n$ denote the nodes at layers $l-1$ and $l$, respectively.

Let the $t$th change of the weight be $\Delta W^l_p(n', s'; n)$ and the $t$th change of the bias be $\Delta b^l(t)$. The error function is defined as

$$E = \frac{1}{2}\,(T - O)^2 \tag{8}$$

where $T$ and $O$ denote the target output value and the actual output value, respectively when the input values $V^0(n', s')$, are

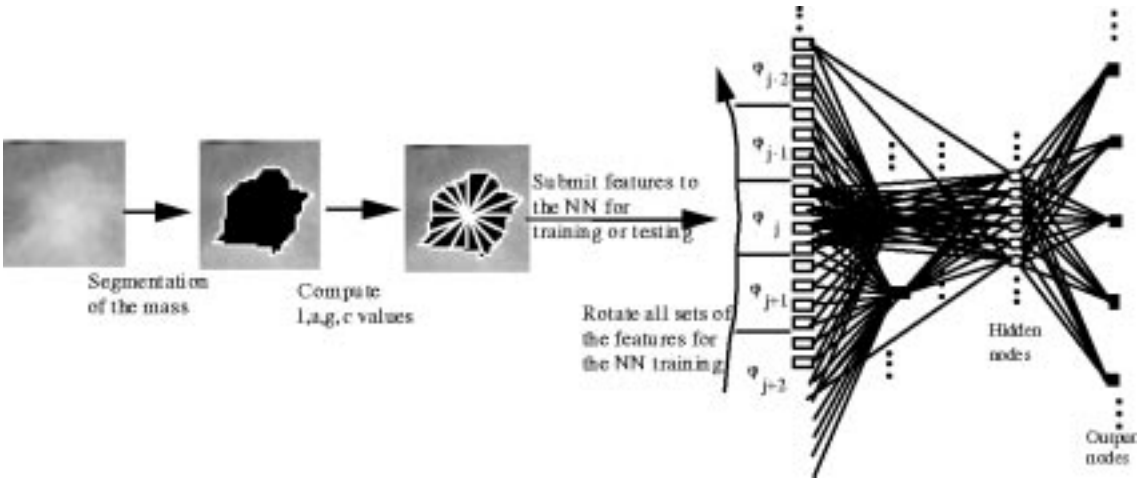


Fig. 5. A schematic diagram, showing the MCPCNN and sector features of masses, that was used in the following study.

entered in the network. In this model, the error backpropagation algorithm, which updates the kernel weights, is given below

$$\Delta W_p^l[t+1] = \eta\left(\sum_n\sum_s \delta_p^{l+1}(n',\, s';n,\, s)\cdot V_p^{l+1}(n,\, s)\right) + \alpha\Delta W_p^l[t] \tag{9}$$

$$\Delta b_p^l[t+1] = \eta\sum_n\sum_s \delta_p^{l+1}(n',\, s';n,\, s) + \alpha\Delta b_p^l[t] \tag{10}$$

$$\delta_p^l(n',\, s';n,\, s) = S'\left(N_p^l(n',\, s')\right)\left(\sum_n\sum_s \delta_p^{l+1}(n,\, s)\cdot W_p^{l+1}(n,\, s)\right). \tag{11}$$

In the case of the last layer

$$\delta^L(n) = S'\left(N^L(n)\right)\left(T(n) - O(n)\right) \tag{12}$$

where $S'(z)$, $\eta$, $\alpha$, and $T$ denote the derivative of $S(z)$, the learning rate, the weighting factor contributed by the momentum term, and the desired output image, respectively. Furthermore, $s$ or $s' = 1$ and $p = 1$ when $l \neq 0$.

During the training, we added an isotropic constraint to the weights of the 1-D convolution kernels so that

$$W_q^0(n, -s) = W_q^0(n,\, s) \tag{13}$$

where $q$ is not the fully connected path. These additional constraints are used to induce the kernels functioning as correlation processing filters and could facilitate the algorithm in searching for an appropriate filter.

*2) Resampling the Training Set Through Utilization of Rotation and Flip Invariance of the Features:* In this neural network model, there are no starting and ending sectors. The forward and backpropagation computation can start from any sector. Considering a flipped patch, the characteristics of mass feature should remain the same. To take advantage of this flip invariance, the same numerical target value can be assigned at the output node for the flipped image patch in order to double the amount of cases during training.

Since we designed a $10^o$ increment for each rotation, every SC or NC path would process through 36 times using the defined features for each image patch. To simplify this network computation, we shifted one small sector (four nodes) on the input layer at a time to conduct the circular convolution process with the SC and NC kernels in the following experiments. By reversing the sequence of the sector, one can train the flipped version of the suspicious masses. Hence, using the properties of the rotation invariance and flip invariance for the neural network training literally increases the number of the training set by a factor of 72.

In summary, we have developed a complete detection procedure for the automatic recognition of mammographic masses including background adjustment, contrast enhancement, ROI segmentation, feature extraction, and MCPCNN system with a training method. Fig. 5 shows a flow diagram for the essential sections of the computational procedures.

## IV. Experiments and Results

As described in Section III-A, the 200 mammograms were selected from the MIAS database and the BAMC database for the study. Of the 200 mammograms, 50 mammograms are normal, and each of the 150 abnormal mammograms contains at least one mass case of varying size, subtlety, and location. Both the cranio-caudal (CC) and medio-lateral oblique (MLO) projection views were used. The films were digitized with a computer format of 2048 $\times$ 2500 $\times$12 bits (for an $8'' \times 10''$ area where each image pixel represents 100 $\mu$m square). Ninety-one mammograms, either a CC or an MLO view film, were selected from 91 patient film jackets. No two mammograms were selected from the same patient. All the digitized mammograms were miniaturized to 512 $\times$ 625 $\times$ 12 bits using $4 \times 4$ pixel averaging before the method was applied. According to radiologists, the size of small masses is 3–15 mm in effective diameter. A 3-mm object in an original mammogram occupies 30 pixels in a digitized image with a 100–$\mu$m resolution. After reducing the image size by four times, the object will occupy the range of about 7–8 pixels. The object with the size of seven pixels

is expected to be detectable by any computer algorithm. After preprocessing and an object screening based on the circularity test and the size test (between 3 and 30 mm), a total of 125 suspicious areas were selected from the testing mammograms (91 cases) for this study. Specifically, the screening procedure of reducing FPs involves two steps: 1) image patches with circularity less than 0.25 or diameter greater than 30 mm were eliminated and 2) ) using probability modular neural network to rule out the majority of FPs. Of the 125 suspicious areas, 75 ROIs contained masses based on corresponding biopsy reports with one experienced radiologist reading. Of 75 masses, 39 were malignant and 36 were benign. This set of ROIs was used in [19] and discussed in [19, Fig. 6 and Table II].

### A. Experiment 1

Of the 125 suspicious areas, we randomly selected 54 computer-segmented ROIs where 30 patches were matched with the radiologist's mass identification and 24 were not. This database was used to train two neural network systems: 1) a conventional three-layer neural network and 2) the proposed MCPCNN training method using the same neural network learning algorithm. The structure of the MCPCNN was described earlier. In the study, we used one fully connected path, four SC paths, four NC paths covering two sectors, four NC paths covering three sectors, three NC paths covering four sectors, and two NC paths covering five sectors in the first step network connection for the MCPCNN. All paths in the neural network have their hidden layers. Only one hidden layer per path was used. Both neural network systems were trained by the error backpropagation algorithm by feeding the features from the input layer and registering the corresponding target value at the output node. Completion of the training was determined by the mean square error [i.e., $\sum_{i=i}^{N}(T_i - O_i)^2/N$, where $N$ is number of samples] when it was approximately reduced to $3 \times 10^{-5}$. Once the training of the neural networks was completed, we then used the remaining 71 computer segmented ROIs for the testing. Forty-five out of 71 ROIs were masses and 26 ROIs were not. Neither the images nor their corresponding patients in the testing set could be found in the training set. The neural network output values were fed into the LABROC4 program [29] for the performance evaluation. The results indicated that the areas ($A_z$) under the receiver operating characteristic (ROC) curves were $0.7869 \pm 0.0536$ and $0.8443 \pm 0.0457$ using the conventional neural network (MLP) and the MCPCNN, respectively. The ROC curves of these two neural network systems are shown in Fig. 6(a). The $A_z$ value was $0.7869 \pm 0.0536$ when using the MLP method with 125 hidden nodes. The performance of the MLP remains about the same at $0.7809 \pm 0.0551$ of $A_z$ using the same neural network parameters but with 30 hidden nodes.

We also invited another senior mammographer to conduct an observer study using the ROC study protocol. The mammographer was asked to rate each patch using a numerical scale rangingfrom zero to ten for its likelihood of being a breast mass. The image patches were displayed on a SUN monitor (Model: GDM-20D10). The image size shown on the monitor was reduced to approximately $7'' \times 9''$ as compared with the original film size ($8'' \times 10''$). These 71 numbers were also fed into the LABROC4 program. The $A_z$ of the mammographer's performance on this set of test cases was $0.909 \pm 0.0340$. The corresponding ROC curve is also shown in Fig. 6(a).

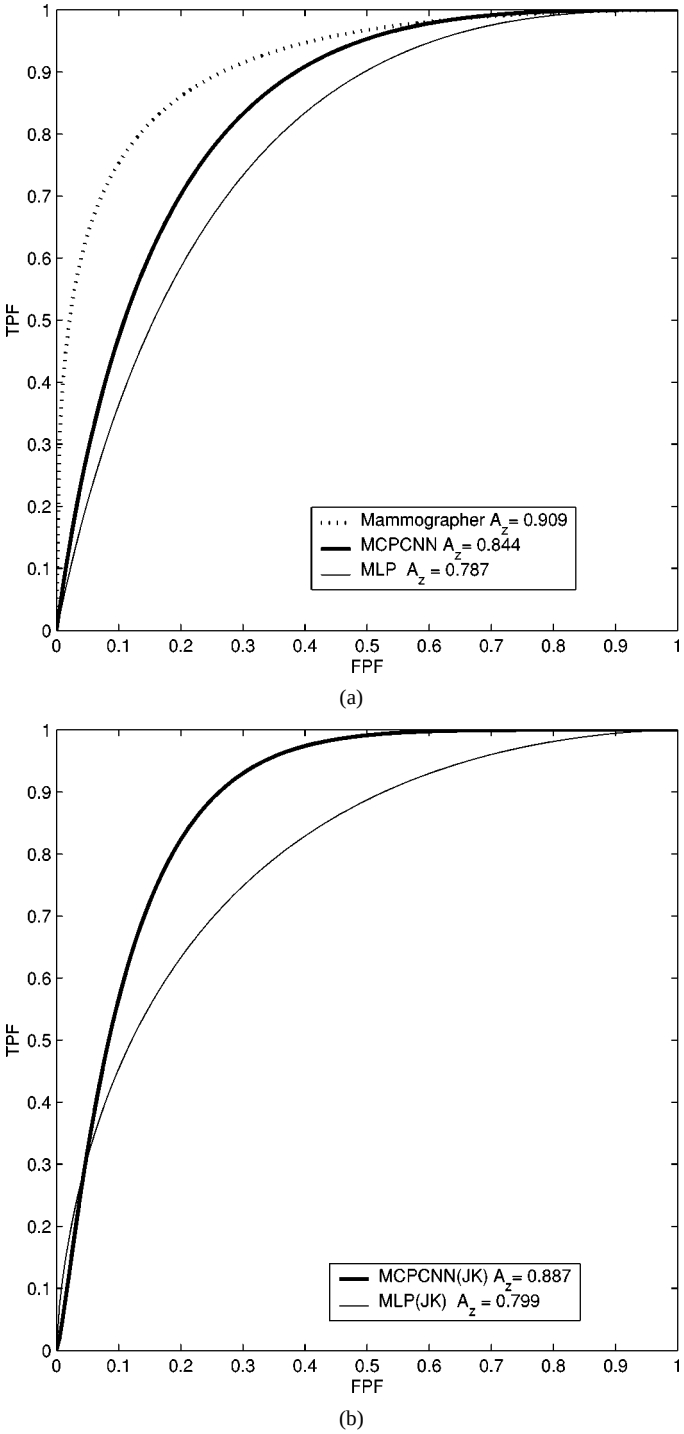


Fig. 6. The ROC curves obtained from corresponding experiments. (a) Shows that the performance of MCPCNN training method is superior to that of the conventional MLP method. The highest curve is the ROC performance of the senior mammographer. (b) Shows that the ROC results were increased using the leave-one-case-out procedure in both neural network systems. The MCPCNN still showed higher performance than conventional MLP method.

### B. Experiment 2

We also conducted a leave-one-case-out experiment (i.e., jackknife procedure) using the same database. In this experiment, we used those image patches extracted from 90

TABLE I
ROC PERFORMANCE OF THE TEST METHODS IN DISTINGUISHING TRUE AND FALSE MASSES

| | Comparative Analyses of Methods | $A_z$ of Method (1) | $A_z$ of Method (2) | P Values | Statistical Significance |
|---|---|---|---|---|---|
| **Experiment 1** | (1) Radiologist vs. (2) MCPCNN | 0.909 ±0.0340 | 0.8443±0.0457 | 0.1855 | No |
| | (1) Radiologist vs. (2) MLP | 0.909 ±0.0340 | 0.7869 ±0.0536 | 0.0447 | **Yes** |
| | (1) MCPCNN vs. (2) MLP | 0.8443±0.0457 | 0.7869 ±0.0536 | 0.1344 | No |
| **Experiment 2** | (1) MCPCNN vs. (2) MLP | 0.8866 ±0.0289 | 0.7985 ±0.0394 | 0.0241 | **Yes** |

mammograms (one mammogram per case) for the training and used the image patches (most of them are single) extracted from the remaining one mammogram as test objects. The procedure was repeated 91 times to allow every ROI extracted from each mammogram to be tested in the experiment. For each individual ROI, the computed features were identical to those used in Experiment 1. Again, the training was stopped when the mean square error value approximately equal to $3 \times 10^{-5}$. Both neural network systems were independently trained and evaluated with the same procedure. The results indicated that the $A_z$ values were $0.7985 \pm 0.0394$ and $0.8866 \pm 0.0289$ using the conventional neural network (MLP) and the MCPCNN, respectively. The performance of the MLP decreased to an $A_z$ of $0.7608 \pm 0.0429$ using the same neural network parameters but with 30 hidden nodes. Fig. 6(b) shows the ROC curves of these two neural network systems using the leave-one-case-out procedure [30] in the experiment.

We also used CLABROC program [31] to analyze the ROC data and compare the ROC results. The results and their statistical significances using two tailed $p$ value of 0.05 as the threshold are shown in Table I. The radiologist's performance is greater than conventional neural network system with a $p$ value of 0.0447 in the first experiment. The MCPCNN was also proven to be superior to the MLP with a statistically significant result ($p = 0.0241$).

## V. DISCUSSION

It is known in the field of artificial intelligence that the key factors in pattern recognition are: 1) effective methods in the extraction of features and 2) classification methods for the extracted features. In this study, we showed that the training method designed to guide the analyzer is also an important factor for a pattern recognition task. Though this finding is not new, the research of developing training methods for various pattern recognition tasks has not been established in the field of medical imaging. Our studies demonstrated that with proper network connections and task-oriented guidance, organized features would assist the neural network in performing the task.

Technically speaking, a feed-forward MLP neural network provides an integrated process for classification and sometimes for feature extraction. The output values of the hidden nodes can be interpreted as a reorganized set of features presented to the output layer for classification. The drawback of the MLP is, the user has a very little control and little understanding about the network learning. The MCPCNN is a network design that partially remedies these issues and is applicable for any pattern recognition task associated with ROIs. The MCPCNN (a member of the CNN family) possesses shared weights in the hidden layer(s) that act as filter kernels for extracting correlated features. With a higher resolution mammogram, a finer sector ($<10°$) would be preferred for the analysis mass, especially for the study of classification of masses. During forward and back-propagation training, the kernels would comply with both signals from input and output layers for all training cases, so as to maximize the classification performance. We do not recommend using 2D CNN for the detection of masses because the mass sizes vary from a few millimeters to 4 cm or even larger. It would require a large fixed size to cover the maximum mass size when using the 2-D CNN. The varieties of mass shapes and potential long spiculated patterns make the use of the 2-D CNN not practical. Since the MCPCNN processes the features computed from sectors, it does not limit the sizes of its ROIs. Best of all, the MCPCNN also has the ability to classify partially obscured masses. The 2-D CNN, however, would be more appropriate for the detection of microcalcifications and small lung nodules.

As far as the research in the detection of masses is concerned, we have shown that use of MCPCNN with sector features is an effective approach. Since the MCPCNN coordinates the input data and performs correlation between features of adjacent sectors in the first stage of data processing, the internal neural network learning algorithm can be changed if a learning algorithm is found to be more effective. In fact, the MCPCNN is a technique that can effectively classify features arranged in the polar coordinate system. A technique using the rubber band straightening transformation, independently developed by Sahnier *et al.* [11], for the detection of masses also employs a similar concept in extracting feature and/or texture in the polar coordinate system. We believe that integration of features and texture values computed at small sectors will be the research trend in mass detection and tumor classification.

## VI. CONCLUSION

In the clinical course of detecting masses, mammographers usually evaluate the surrounding background of a radiodense area when an ROI is suspected. In this study, we simulated this fundamental concept with a neural network system (i.e., MCPCNN). In order for the MCPCNN to function, boundary features of the suspicious region in each radial sector were computed. We found that the MCPCNN is capable of analyzing correlated features within the sector and between adjacent sectors, which led to an improvement in detecting mammographic masses.

Through this study, we found that the selected features are somewhat effective in the detection of masses. These features

were "computationally translated" from the qualitative descriptors of BI—RAD. These features can be extended for the improvement of the mass detection, but this task is beyond the scope of this paper. With the preliminary studies shown above, we found the MCPCNN coupling with the proposed training method produced greater results than the conventional neural network. We found that the performances of both neural network systems were improved in Experiment 2. This may have occurred due to the number of training samples that was increased from 54 to 124. In Experiment 2, the $A_z$ value was improved by 0.042 using the MCPCNN, which was higher than the $A_z$ difference of 0.012 obtained by the conventional training method. The results implied that the MCPCNN learned more effectively than the conventional neural network when the number of training cases was increased. With the use of a larger database and advanced texture features proposed by others, it is expected that the performance of MCPCNN should be significantly improved. This paper does not intend to claim the best mass detection system, in comparison to similar systems; but rather its goal is to report a potentially better neural network structure for analyzing a set of mass features.

## ACKNOWLEDGMENT

A part of the database, used in the study, was provided by Dr. R. Shah of Brooke Army Medical Center. The LABROC4 and CLABROC programs were written by Dr. C. E. Metz and his colleagues at the University of Chicago.

## REFERENCES

[1] L. Nystrom, L. E. Rutqvist, S. Wall, A. Lindgren, M. Lindqvist, and S. Ryden *et al.*, "Breast cancer screening with mammography: Overview of Swedish randomized trials," *Lancet*, vol. 341, pp. 973–978, 1993.

[2] S. Shapiro, "Screening-assessment of current studies," *Cancer*, vol. 74, pp. 231–238, 1994.

[3] L. Tabar, G. Fagerberg, S. Duffy, N. E. Day, A. Gad, and O. Grontoft, "Update of the Swedish two-country program of mammographic screening for breast cancer," *Radiol. Clin. N. Amer.: Breast Imag.—Current Status Future Directions*, vol. 30, pp. 187–210, 1992.

[4] D. Brzakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammograms," *IEEE Trans. Med. Imag.*, vol. 9, p. 233, Sept. 1990.

[5] R. Zwiggelaar, T. C. Parr, J. E. Schumm, I. W. Hutt, C. J. Taylor, S. M. Astley, and C. R. M. Boggis, "Model-based detection of spiculated lesions in mammograms," *Med. Image Anal.*, vol. 3, no. 1, pp. 39–62, 1999.

[6] N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," *Med. Phys.*, vol. 23, no. 10, pp. 1685–1696, 1996.

[7] B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissues: A convolution neural network classifier with spatial domain and texture images," *IEEE Trans. Med. Imag.*, vol. 15, pp. 598–610, Oct. 1996.

[8] D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograms: Multiresolution texture analysis," *Med. Phys.*, vol. 25, no. 4, pp. 516–526, 1998.

[9] L. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, and M. A. Helvie, "Classification of malignant and benign masses based on hybrid ART2LDA approach," *IEEE Trans. Med. Imag.*, vol. 18, pp. 1178–1187, Dec. 1999.

[10] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized detection of malignant tumors on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 18, pp. 369–378, May 1999.

[11] B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and textures analysis," *Med. Phys.*, vol. 25, no. 4, pp. 516–526, 1998.

[12] M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 17, pp. 510–517, Aug. 1998.

[13] D. D. Adler, "Breast Masses: Differential Diagnosis," in *ARRS Categorical Course Syllabus on Breast Imaging*, S. A. Feig, Ed. Reston, VA: Amer. Roent. Ray Soc., 1988, p. 31.

[14] M. J. Homer, "Imaging features and management of characteristically benign and probably benign breast lesions," *Radiol. Clin. N. Amer.*, vol. 25, p. 939, 1987.

[15] S. Pohlman, K. A. Powell, N. A. Obuchowski, W. A. Chilcote, and S. Grundfest-Broniatowski, "Quantitative classification of breast tumors in digitized mammograms," *Med. Phys.*, vol. 23, no. 8, pp. 1337–1345, 1996.

[16] M. Moskowitz, "Circumscribed lesions of the breast," in *Diagnostic Categorical Course in Breast Imaging*, M. Moskowitz, Ed. Oak Brook, IL: Radiol. Soc. N. Amer., 1986, p. 31.

[17] E. A. Sickles, "The rule of multiplicity and the developing density sign," in *ARRS Categorical Course Syllabus on Breast Imaging*, S. A. Feig, Ed. Reston, VA: Amer. Roent. Ray Soc., 1988, p. 177.

[18] H. Li, Y. Wang, K-J. R. Liu, S-C. B. Lo, and M. T. Freedman, "Computerized radiographic mass detection—Part I: Lesion site selection by morphological enhancement and contextual segmentation," *IEEE Trans. Med. Imag.*, pp. 289–301, Apr. 2001.

[19] ——, "Computerized radiographic mass detection—Part II: Decision support by featured database visualization and modular neural networks," *IEEE Trans. Med. Imag.*, pp. 302–313, Apr. 2001.

[20] *Breast Imaging—Reporting and Data System*. Reston, VA: Ame. Coll. Radiol., 1993.

[21] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, "The mammographic images analysis society digital mammogram database," in *Exerpta Medica*, ser. Int. Congr., 1994, vol. 1069, (e-mail for inquiry: mias@sv1.smb.man.ac.uk.), pp. 375–378.

[22] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall., 1999.

[23] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representation by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA: M.I.T. Press, 1986, vol. 1, Foundation, ch. 8, pp. 318–362.

[24] S.-C. B. Lo, S. L. Lou, J. S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications to lung nodule detection," *IEEE Trans. Med. Imag.*, vol. 14, pp. 711–718, Dec. 1995.

[25] S.-C. B. Lo, H. P. Chan, J. S. Lin, H. Li, M. T. Freedman, and S. K. Mun, "Artificial convolution neural network for medical image pattern recognition," *Neural Networks*, vol. 8, no. 7/8, pp. 1201–1214, 1995.

[26] H. P. Chan, S.-C. B. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided diagnosis of mammographic microcalcifications: Pattern recognition with an artificial neural network," *Med. Phys.*, vol. 24, no. 10, pp. 1555–1567, 1995.

[27] Y. Wu, K. Doi, M. L. Giger, and R. M. Nishikawa, "Computerized detection of clustered microcalcifications in digital mammograms: Applications of artificial neural networks," *Med. Phys.*, vol. 19, pp. 555–560, 1992.

[28] Y. Wu, M. T. Freedman, S.-C. B. Lo, R. A. Zuurbier, A. Hasegawa, and S. K. Mun, "Classification of microcalcifications in radiographs of pathological specimen for the diagnosis of breast cancer," *Acad. Radiol.*, vol. 2, pp. 199–204, 1995.

[29] C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Statist. Med.*, vol. 17, pp. 1033–1053, 1998.

[30] K. Fukunaga and R. R. Hayes, "Effects of sample size in classifier design," *IEEE Trans. Pattern Anal. Machine Intell.*, pp. 873–885, Aug. 1989.

[31] C. E. Metz, P.-L. Wang, and H. B. Kronman, "A new approach for testing the significance of differences between ROC curves measured from correlated data," in *Information Processing in Medical Imaging*, F. Deconinck, Ed. The Hague, The Netherlands: Martinus Nijhoff, 1984, vol. PAMI-II, pp. 432–445.