

Computer-Aided Diagnostic Scheme for Distinction Between Benign and Malignant Nodules in Thoracic Low-Dose CT by Use of Massive Training Artificial Neural Network

Kenji Suzuki*, *Senior Member, IEEE*, Feng Li, Shusuke Sone, and Kunio Doi

Abstract—Low-dose helical computed tomography (LDCT) is being applied as a modality for lung cancer screening. It may be difficult, however, for radiologists to distinguish malignant from benign nodules in LDCT. Our purpose in this study was to develop a computer-aided diagnostic (CAD) scheme for distinction between benign and malignant nodules in LDCT scans by use of a massive training artificial neural network (MTANN). The MTANN is a trainable, highly nonlinear filter based on an artificial neural network. To distinguish malignant nodules from six different types of benign nodules, we developed multiple MTANNs (multi-MTANN) consisting of six expert MTANNs that are arranged in parallel. Each of the MTANNs was trained by use of input CT images and teaching images containing the estimate of the distribution for the “likelihood of being a malignant nodule,” i.e., the teaching image for a malignant nodule contains a two-dimensional Gaussian distribution and that for a benign nodule contains zero. Each MTANN was trained independently with ten typical malignant nodules and ten benign nodules from each of the six types. The outputs of the six MTANNs were combined by use of an integration ANN such that the six types of benign nodules could be distinguished from malignant nodules. After training of the integration ANN, our scheme provided a value related to the “likelihood of malignancy” of a nodule, i.e., a higher value indicates a malignant nodule, and a lower value indicates a benign nodule. Our database consisted of 76 primary lung cancers in 73 patients and 413 benign nodules in 342 patients, which were obtained from a lung cancer screening program on 7847 screenees with LDCT for three years in Nagano, Japan. The performance of our scheme for distinction between benign and malignant nodules was evaluated by use of receiver operating characteristic (ROC) analysis. Our scheme achieved an Az (area under the ROC curve) value of 0.882 in a round-robin test. Our scheme correctly identified 100% (76/76) of malignant nodules as malignant, whereas 48% (200/413) of benign nodules were identified correctly as benign. Therefore, our scheme may be useful in assisting radiologists in the diagnosis of lung nodules in LDCT.

Index Terms—Artificial neural network, computer-aided diagnosis (CAD), likelihood of malignancy, low-dose CT, lung nodule.

Manuscript received July 30, 2004; revised May 17, 2005. This work was supported by the USPHS under Grant CA62625. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was H.-P. Chan. *Asterisk indicates corresponding author.*

*K. Suzuki is with the Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637 USA (e-mail: suzuki@uchicago.edu).

F. Li and K. Doi are with the Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, The University of Chicago, Chicago, IL 60637 USA.

S. Sone is with Azumi General Hospital, Ikeda, Kitaazumi, Nagano, 399-8695, Japan

Digital Object Identifier 10.1109/TMI.2005.852048

I. INTRODUCTION

LUNG CANCER continues to rank as the leading cause of cancer deaths among Americans; the number of lung cancer deaths in each year is greater than the combined number of breast, colon, and prostate cancer deaths [1]. Because CT is more sensitive than chest radiography in the detection of small nodules and of lung carcinoma at an early stage [2]–[4], lung cancer screening programs are being investigated in the United States [2], [5]–[10] and Japan [3], [11]–[13] with low-dose helical computed tomography (LDCT) as the screening modality. It may be difficult, however, for radiologists to distinguish between benign and malignant nodules on LDCT. In a screening program with LDCT in New York, 88% (206/233) of suspicious lesions were found to be benign nodules on follow-up examinations [5]. In a screening program in Japan, only 83 (10%) among 819 scans with suspicious lesions were diagnosed to be cancer cases [13]. According to recent findings at the Mayo Clinic, 2 792 (98.6%) of 2 832 nodules detected by a multidetector CT were benign, and 40 (1.4%) nodules were malignant [7]. Thus, a large number of benign nodules were found with CT; follow-up examinations such as high-resolution CT (HRCT) and/or biopsy were performed on these patients. Therefore, computer-aided diagnostic (CAD) schemes for distinction between benign and malignant nodules in LDCT would be useful for reducing the number of “unnecessary” follow-up examinations.

Our purpose in this study was to develop a CAD scheme for distinction between benign and malignant nodules in LDCT by use of a new pattern-classification technique based on a massive training artificial neural network (MTANN).

II. MATERIALS

Our database consisted of 76 primary lung cancers in 73 patients and 413 benign nodules in 342 patients, which were obtained from a lung cancer screening program on 7 847 screenees with LDCT from 1996 to 1999 in Nagano, Japan [4]. All cancers were confirmed histopathologically at either surgery or biopsy. During the initial clinical reading, all benign nodules were reported as lesions suspected to be lung cancer or indeterminate lung lesions, but were not reported as benign cases. The CT examinations were performed on a mobile CT scanner (CT-W950SR; Hitachi Medical, Tokyo, Japan). The scans used for this study were acquired with a low-dose protocol of 120 kVp,

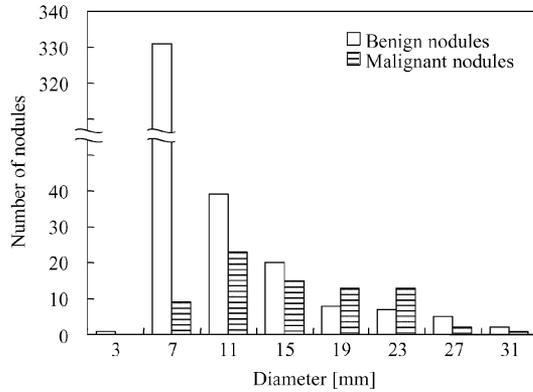


Fig. 1. Distributions of sizes of malignant and benign nodules in our database.

25 mA or 50 mA, 10-mm collimation, and a 10-mm reconstruction interval at a helical pitch of two. The pixel size was 0.586 mm or 0.684 mm. Each reconstructed CT section had an image matrix size of 512×512 pixels. The nodule size ranged from 3 mm to 29 mm. When a nodule was present in more than one section, the section with the greatest size as determined by an experienced chest radiologist was used in this study, because either a computerized detection scheme for lung nodules or a radiologist is likely to select the section with the greatest nodule size or a similar size rather than the section with a small nodule size where the nodule may appear to be small and of low contrast. The centers of nodules were identified by an experienced chest radiologist, and were used for extracting regions of interest (ROIs) for training and testing of our scheme. The 76 primary lung cancers consisted of 22 (28.9%) nodules identifiable in a single section, 37 (48.7%) nodules in two sections, and 17 (22.3%) nodules in three sections. The 413 benign nodules consisted of 265 (64.2%) nodules in a single section, 133 (32.2%) in two sections, and 15 (3.6%) nodules in three sections. Figure 1 shows the histograms of sizes of malignant and benign nodules in our database. Approximately 30% of lung cancers were attached to the pleura, 34% of cancers were attached to vessels, and 7% of cancers were in the hilum. Three chest radiologists independently reviewed the cancers, and then classified them in three categories, pure ground-glass opacity (pure GGO; 24% of cancers), mixed GGO (30%), and solid nodule (46%) by consensus. Thus, this database included various types of nodules of various sizes.

III. METHOD

A. Architecture of Massive Training Artificial Neural Network (MTANN)

Suzuki *et al.* have been investigating supervised nonlinear image-processing techniques based on artificial neural networks (ANNs), called a “neural filter” [14], for reduction of the quantum mottle in x-ray images [15] and a “neural edge detector” [16], [17] for supervised detection of subjective edges traced by cardiologists [18], and they have developed training methods [19], [20], design methods [21]–[23], and an analysis method [24] for these techniques. Suzuki *et al.* recently extended the neural filter and the neural edge detector to accommodate various pattern-classification tasks, and they

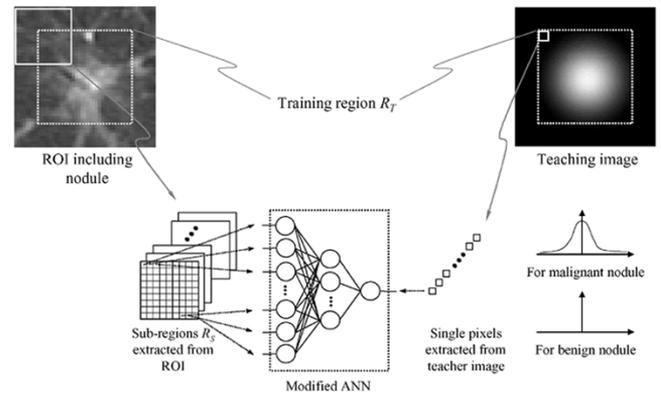


Fig. 2. Architecture and training of a massive training artificial neural network (MTANN) for distinction between benign and malignant nodules. The pixel values in the subregions extracted from the region of interest (ROI) are entered as input to the MTANN. The single pixel corresponding to the input subregion, which is extracted from the teaching image, is used as a teaching value.

developed an MTANN. They have applied the MTANN for reduction of false positives in computerized detection of lung nodules in LDCT [25]–[28] and chest radiography [29].

The architecture and the training method of the MTANN are shown in Fig. 2. The MTANN is a highly nonlinear filter that can be trained by use of input images and the corresponding teaching images. The MTANN consists of a modified multi-layer ANN in which layers are fully connected with adjustable weights [30], and which is capable of operating on image data directly. The MTANN employs a linear function instead of a sigmoid function as the activation function of the unit in the output layer because the characteristics of an ANN were significantly improved with a linear function when applied to the continuous mapping of values in image processing, [17] for example. Note that the activation functions of the units in the hidden layer are a sigmoid function for nonlinear processing, and those of the units in the input layer are an identity function, as usual. The pixel values of the original CT images are normalized first such that -1000 HU (Hounsfield units) is zero and 1000 HU is one. The inputs of the MTANN are the pixel values in a local window R_S on a region of interest (ROI) in a CT image. The output of the MTANN is a continuous value, which corresponds to the center pixel in the local window, represented by

$$O(x, y) = \text{NN}\{I(x - i, y - j) \mid i, j \in R_S\} \quad (1)$$

where $O(x, y)$ is the output of the MTANN, x and y are the indices of coordinates, $\text{NN}\{\cdot\}$ is the output of the modified multilayer ANN, and $I(x, y)$ is an input pixel value. Note that only one unit is employed in the output layer. The output image is obtained by scanning of an input image with the MTANN. The local window of the MTANN must be shifted pixel by pixel throughout the input image.

B. Training of MTANN

For distinguishing malignant nodules from benign nodules, the teaching image contains the estimate of the distribution for the “likelihood of being a malignant nodule,” i.e., the teaching image for a malignant nodule should contain a certain distribution, the peak of which is located at the center of the malignant

nodule, and that for a benign nodule should contain zeros. As the distance from the center of the malignant nodule increases, its likelihood of being a malignant nodule decreases; therefore, we used a two-dimensional (2-D) Gaussian function with standard deviation σ_T at the center of the malignant nodule as the distribution for the likelihood of being a malignant nodule, where σ_T may be determined as a measure representing the size of malignant nodules. Figure 2 illustrates the training for an input image that contains a malignant nodule near the center. The training region R_T in the input image is divided pixel by pixel into a large number of overlapping subregions, the size of which corresponds to that of the local window R_S of the MTANN. The centers of consecutive subregions in Fig. 2 differ by just one pixel. All pixel values in each of the subregions are entered as input to the MTANN, whereas one pixel from the teaching image is entered into the output unit in the MTANN as the teaching value. This single pixel is chosen at the location in the teaching image that corresponds to the center of the input subregion. Thus, the MTANN is trained by presenting each of the input subregions together with each of the corresponding teaching single pixels. A large number of input subregions overlap each other, and the corresponding teaching pixels in the teaching image are used for training. The MTANN is trained with massive training samples to achieve a high generalization ability. The MTANN would be robust against variation in patterns, especially shifted patterns, because it is trained with numerous shifted patterns. The MTANN would be able to learn the essential features of nodules without dependence on spatial shift. The error to be minimized by training is defined by

$$E = \frac{1}{N_s \cdot P} \sum_{s=1}^{N_s} \sum_{x,y \in R_T} \{T_s(x,y) - O_s(x,y)\}^2 \quad (2)$$

where $T_s(x,y)$ is the teaching image for the s th training ROI (a malignant nodule or a benign nodule), $O_s(x,y)$ is the output image for the s th training ROI, R_T is the training region, N_s is the number of training ROIs, and P is the number of training pixels in R_T . The MTANN is trained by a modified back-propagation (BP) algorithm [30], which was derived for the modified multilayer ANN, i.e., a linear function is employed as the activation function of the unit in the output layer, in the same way as the original BP algorithm [31], [32]. The MTANN is trained by adjustment of the weights between layers iteratively so that the error becomes small. After training, the MTANN is expected to output the highest value when a malignant nodule is located at the center of the local window of the MTANN, a lower value as the distance from the center increases, and zero when the input region contains a benign nodule.

C. Multiple MTANNs

In order to distinguish malignant nodules from various types of benign nodules, we extended the capability of a single MTANN and developed multiple MTANNs (multi-MTANN) [25]. The architecture of the multi-MTANN is shown in Fig. 3. The multi-MTANN consists of plural MTANNs that are arranged in parallel. Each MTANN is trained by use of benign nodules representing a different benign type, but with the same malignant nodules. Each MTANN acts as an expert for

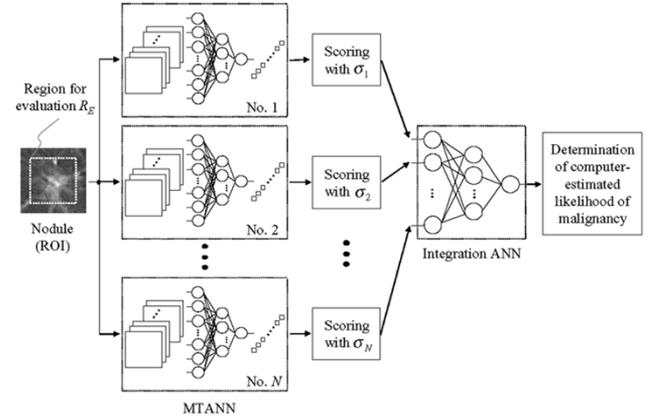


Fig. 3. Architecture of multiple MTANNs (multi-MTANN) incorporating an integration artificial neural network (ANN) for distinguishing malignant nodules from various benign nodules. Each MTANN is trained by use of benign nodules representing a different benign type, but with the same malignant nodules. Each MTANN acts as an expert for distinction between malignant nodules and a specific type of benign nodule. The output of each MTANN is integrated by use of the integration ANN.

distinguishing malignant nodules from a specific type of benign nodule, e.g., MTANN no. 1 is trained to distinguish malignant nodules from small benign nodules overlapping with vessels; MTANN no. 2 is trained to distinguish malignant nodules from medium-sized benign nodules with fuzzy edges; and so on.

The distinction between a malignant nodule and a benign nodule is determined by use of a score defined from the output image of the trained n th MTANN, represented by

$$S_n = \sum_{x,y \in R_E} f_G(\sigma_n; x,y) \times O_n(x,y) \quad (3)$$

where S_n is the output score for a given nodule from the n th MTANN, R_E is the region for evaluation, $O_n(x,y)$ is the output image of the n th MTANN where its center corresponds to the center of R_E , and $f_G(\sigma_n; x,y)$ is a 2-D Gaussian weighting function with standard deviation σ_n , where its center corresponds to the center of R_E . This score represents the weighted sum of the estimate for the likelihood that the image contains a malignant nodule near the center, i.e., a higher score would indicate a malignant nodule, and a lower score would indicate a benign nodule. The concept of this scoring is similar to that of the matched filter. We used a 2-D Gaussian weighting function, because this function should be the same one used in the teaching images, which was a 2-D Gaussian function. The function in the teaching images represents the estimate of the distribution of a “likelihood of being a malignant nodule.” The function in the teaching images can be changed by following this concept. Consequently, the weighting function for scoring should be changed to the same function in the teaching images.

It is difficult to distinguish a small output distribution for a small malignant nodule from a small distribution due to noise. This can lower the performance in distinguishing malignant nodules from benign nodules. We used the same-sized Gaussian distribution in the teaching images, because we intended to force the MTANN to output a regular-sized distribution for different-sized nodules, e.g., a larger output distribution for a small nodule. After training in this way, the MTANN expects to

output relatively regular-sized distributions for different-sized malignant nodules, e.g., a relatively larger output distribution for a small nodule and a relatively smaller output distribution for a large nodule. This property of the regular-sized output distributions expects to make the scores for small malignant nodules higher, and to contribute to improvement of the overall performance of the MTANN.

D. Integration ANN

The scores from the expert MTANNs in the multi-MTANN are combined by use of an integration ANN such that different types of benign nodules can be distinguished from malignant nodules. The integration ANN consists of a modified multilayer ANN with a modified BP training algorithm [30] for processing continuous output/teaching values, i.e., the activation functions of the units in the input, hidden, and output layers are an identity, a sigmoid, and a linear function, respectively. The layers are fully connected with adjustable weights. The scores of each MTANN are entered to each input unit in the integration ANN; thus, the number of input units corresponds to the number of MTANNs. The scores of each MTANN function like the features for distinguishing malignant nodules from a specific type of benign nodule with which the MTANN was trained. One unit is employed in the output layer for distinction between a malignant nodule and a benign nodule. The teaching values for malignant nodules are assigned the value one, and those for benign nodules are zero. After training, the integration ANN is expected to output a higher value for a malignant nodule, and a lower value for a benign nodule. Thus, the output can be considered to be a value related to a “likelihood of malignancy” of a nodule. By thresholding of the output, a distinction between malignant and benign nodules can be made. The balance between a true-positive rate (TPR) and a false-positive rate (FPR) is a choice, which can be determined by the threshold value. If the scores of each MTANN characterize the specific type of benign nodule with which the MTANN is trained, then the integration ANN combining several MTANNs will be able to distinguish malignant nodules from various types of benign nodules.

IV. RESULTS

A. Training

For selecting the training malignant nodules for an MTANN, we classified malignant nodules into several groups based on the visual appearance of patterns in terms of size, solidity (solid or nonsolid), spiculation, contrast, and background. We selected one or two nodules from each group, and obtained ten typical malignant nodules. We selected ten small benign nodules with vessels as training benign nodules, because these nodules were dominant over all benign nodules in our database. We trained the MTANN with the ten malignant nodules and ten benign nodules (the parameters for the MTANN are described latter). Then we applied the trained MTANN to the entire database to obtain scores for all nodules. For selecting training benign nodules for MTANNs in a multi-MTANN, we classified benign nodules into seven other groups by using a method for determining training cases for a multi-MTANN [26]. With this method, training cases

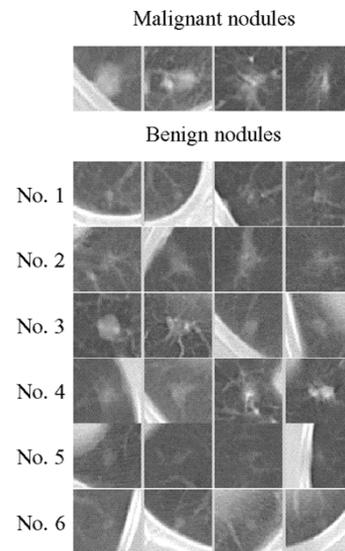


Fig. 4. Illustrations of training samples of four malignant nodules (top row) and six sets of four benign nodules for six MTANNs in the multi-MTANN.

for each MTANN were determined systematically based on the ranking in the scores obtained from the first trained MTANN so that benign nodules in each group cause different degrees of difficulty in classification by the MTANN. We selected ten benign nodules from each of the groups. We used six out of eight groups (the first group plus the seven groups) as training cases for the multi-MTANN by experimental analysis (described in Section V). Figure 4 shows samples of training cases for malignant and benign nodules. The six groups included 1) small nodules overlapping with vessels, 2) medium-sized nodules with fuzzy edges, 3) medium-sized nodules with sharp edges and relatively small nodules with light background, 4) medium-sized nodules with high contrast and medium-sized nodules with light background, 5) small nodules with fuzzy edges, and 6) small nodules near the pleura. A three-layer structure was employed as the structure of the MTANN, because any continuous mapping can be realized approximately by three-layer ANNs [33], [34]. The size of the local window R_S of the MTANN, the standard deviation σ_T of the 2-D Gaussian function, and the size of the training region R_T in the teaching image were determined to be 9×9 pixels, 5.0 pixels, and 19×19 pixels, respectively, by use of the results of the experimental analysis described in [25]. The performance of the MTANN was the highest when these values were used. These parameters were fixed, and the same parameters were used for all six MTANNs. The number of hidden units was determined to be 20 units by use of the results of the experimental analysis described in [25]. Thus, the numbers of units in the input, hidden, and output layers were 81, 20, and 1, respectively. The slope of the linear function of the output units of MTANNs, and the learning rate for training the MTANNs, were 0.01 and 0.002, respectively. With the parameters above, the training of each MTANN in the multi-MTANN was performed 500 000 times. The training of each MTANN required a CPU time of 29.8 h on a PC-based workstation (CPU: Pentium IV, 1.7 GHz). Before we applied the trained MTANN, pixels outside the segmented lung regions [54] were set to -1000 HU in order to reduce the effect of strong edges of the pleura. The

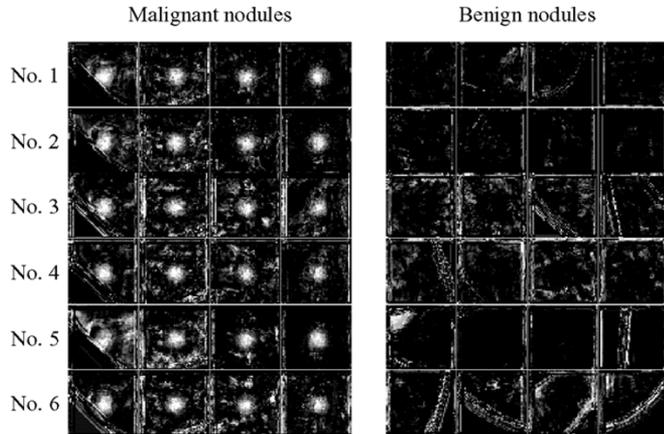


Fig. 5. Illustrations of the output images of the six trained MTANNs for malignant nodules (left four images) and benign nodules (right four images), which correspond to the training samples in Fig. 4. Note that the output images of each MTANN for malignant nodules correspond to the same four input images.

output images of each trained MTANN for training cases are shown in Fig. 5.

We applied the scoring method to the output image of the MTANNs. The standard deviation σ_n of the 2-D Gaussian weighting function was changed from 5.0 to 10.0 in steps of 0.5, and was determined to be the standard deviation with the highest Az value for each MTANN. As a result, the standard deviations were determined to be 7.5 or 8.0. The size of the evaluation region R_E was 19×19 pixels, which was the same size as the training region R_T . The scores of each trained MTANN in the multi-MTANN were used as inputs to the integration ANN with a three-layer structure. The number of hidden units in the integration ANN was determined to be four by experimental analysis (described in the Discussion section). Thus, the numbers of units in the input, hidden, and output layers were six, four, and one, respectively. The slope of the linear function of the output unit and the learning rate were 0.1 and 0.1, respectively. The training of the integration ANN was performed 1 000 times with the round-robin (leave-one-out) test. With this test, one nodule was excluded from all nodules, and the remaining nodules were used for training of the integration ANN. After training, the one nodule excluded from training cases was used for testing. This process was repeated for each of the nodules one by one, until all nodules were tested.

B. Evaluation

The trained MTANNs in the multi-MTANN were applied to our database of 76 malignant nodules and 413 benign nodules. Figure 6 shows input images and the corresponding output images of each of the six MTANNs for nontraining cases. The malignant nodules in the output images of the MTANN were represented by light distributions near the centers of the nodules, whereas the benign nodules in the corresponding group for which the MTANN was trained in the output images were mostly dark around the center, as expected. Figure 7 shows nontraining malignant nodules representing three major types of patterns, i.e., pure GGO, mixed GGO, and solid nodule, and the corresponding output images of the MTANN no. 1 for distinction of malignant from benign nodules in the group (1).

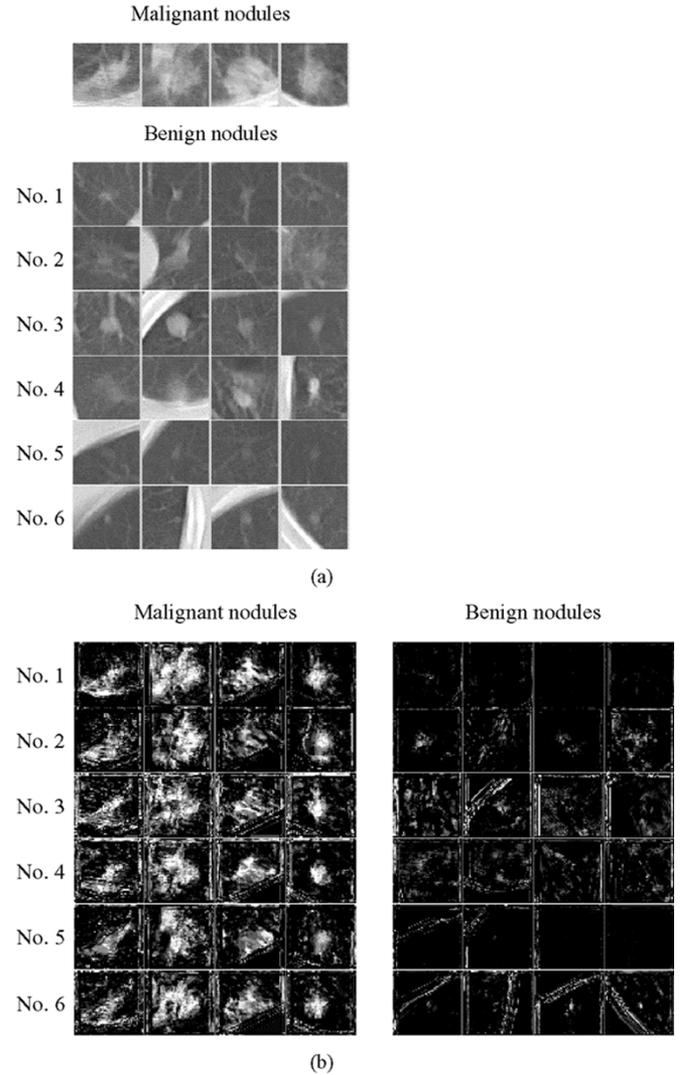


Fig. 6. Illustrations of (a) four nontraining malignant nodules (top row) and six nontraining sets of four benign nodules, and (b) the corresponding output images of the six trained MTANNs in the multi-MTANN for malignant nodules (left four images) and benign nodules (right four images).

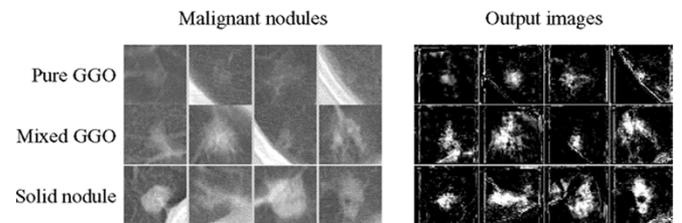


Fig. 7. Illustration of three types of nodule patterns, i.e., pure GGO, mixed GGO, and solid nodule, and the corresponding output images of the trained MTANN no. 1 for nontraining cases.

All three types of nodules are represented by light distributions. The distributions are relatively regular-sized for different-sized malignant nodules, e.g., a relatively larger output distribution for a small nodule and a relatively smaller output distribution for a large nodule. The scoring method was applied to the output images. The performance of each MTANN was evaluated by receiver operating characteristic (ROC) analysis [35], [36]. Figure 8 shows the ROC curve of each MTANN for nontraining cases of 66 malignant nodules and 403 benign nodules.

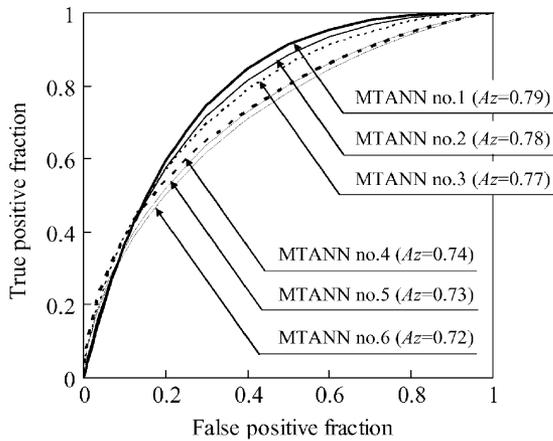


Fig. 8. ROC curve of each MTANN in the multi-MTANN in distinction between 66 nontraining malignant nodules and 403 nontraining benign nodules.

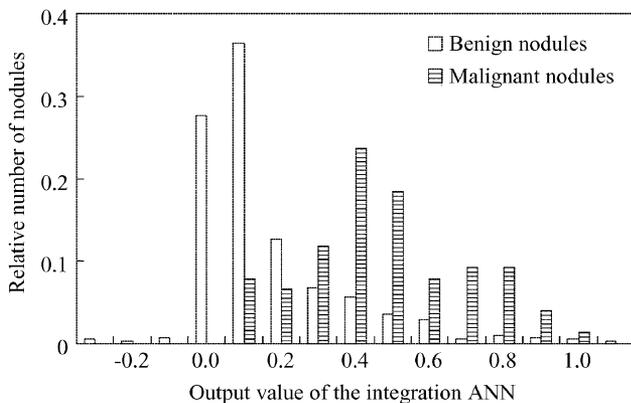


Fig. 9. Distributions of the output values of the integration ANN for 76 malignant nodules and 413 benign nodules in the round-robin test.

Although the A_z values did not differ very much, the scores from each MTANN appropriately characterized a specific type of benign nodule with which the MTANN was trained, i.e., the scores from the MTANN for the corresponding type of benign nodule were low, whereas those for malignant nodules were substantially high. It is very important for combining the MTANNs that the characteristics of each MTANN differ. If each MTANN in the multi-MTANN has exactly the same characteristics, the performance of our scheme would not be improved by combining them.

Figure 9 shows the distributions of the output values of the trained integration ANN for the 76 malignant nodules and 413 benign nodules in the round-robin test. Although the two distributions overlap, malignant nodules can be distinguished from some benign nodules. Table I illustrates TPRs and the corresponding FPRs obtained by thresholding of the output values with different threshold levels. Our scheme achieved a TPR of 100% with an FPR of 51.6% at a certain threshold level. By changing threshold levels, we obtained pairs of TPR and FPR: 94.7% with 35.4%, 90.8% with 33.4%, and 80.0% with 15.7%. The performance of our scheme based on the multi-MTANN incorporated with the integration ANN was evaluated by ROC analysis [35], [36]. We used the output values from the integration ANN as scores in the ROC analysis. Binormal distributions were fitted to the scores by use of maximum-likelihood estimation [36]. The ROC curve was obtained by changing the

TABLE I
TPRS AND THE CORRESPONDING FPRS OF OUR SCHEME AT DIFFERENT THRESHOLD LEVELS

TPR	FPR
1.000	0.516
0.987	0.477
0.974	0.397
0.961	0.390
0.947	0.354
0.934	0.351
0.921	0.341
0.908	0.334
0.800	0.157
0.700	0.140
0.600	0.104
0.500	0.087
0.400	0.073
0.300	0.036
0.200	0.029
0.100	0.017
0.000	0.005

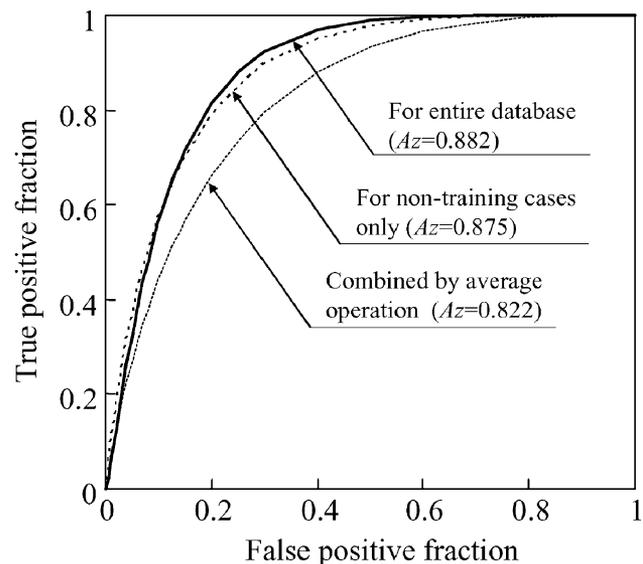


Fig. 10. ROC curves of our schemes in distinction between malignant and benign nodules. The solid curve indicates the performance (A_z value of 0.882) of our scheme in distinction between 76 malignant nodules and 413 benign nodules in the round-robin test. The performance is higher at high sensitivity levels. The dashed curve indicates the performance (A_z value of 0.875) of our scheme for nontraining cases of 66 malignant nodules and 353 benign nodules. The dotted curve indicates the performance (A_z value of 0.822) of the multi-MTANN, the outputs of which were combined with the average operation.

threshold value (decision variable), and represented true-positive fractions as a function of false-positive fractions. Figure 10 shows the ROC curve of our scheme. This scheme achieved an A_z value (area under the ROC curve) [37] of 0.882 (standard error = 0.0167) in the round-robin test. The performance for nontraining cases, i.e., the training cases of ten malignant nodules and 60 benign nodules were excluded from the cases for evaluation, was almost the same (A_z value of 0.875). The ROC curve was higher at high sensitivity levels. This allows us to distinguish many benign nodules without loss of a malignant nodule. Our scheme correctly identified 100% (76/76) of malignant nodules as malignant, and 48% (200/413) of benign nodules were identified correctly as benign.

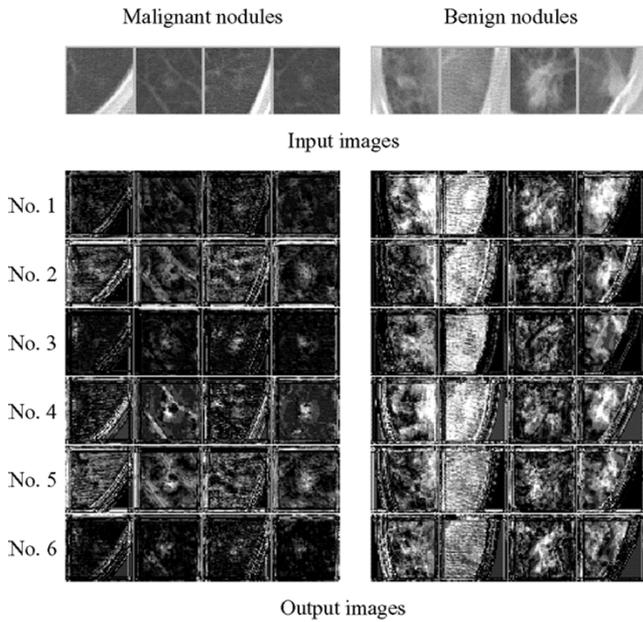


Fig. 11. Four worst cases of malignant and benign nodules for our scheme and the corresponding output images of the MTANNs in the multi-MTANN.

Figure 11 shows malignant and benign nodules with the four worst scores, i.e., malignant nodules with the four lowest scores and benign nodules with the four highest scores, which correspond to the output values at the left-most side of the distribution for malignant nodules and those at the right-most side of the distribution for benign nodules in Fig. 9, respectively. The malignant nodules with the four worst scores were relatively small, and had no remarkable malignant features in their appearance. In the output image of the majority of the MTANNs, these small malignant nodules are dark. The benign nodules with the four worst scores were relatively small nodules with a light background, which correspond to one category in the third group of the training cases, and relatively large nodules with spiculation. The MTANNs could not output lower values for the light background, probably because the background of these nodules was lighter than that of the training cases in the third group, as shown in the third and fourth ROIs in the third row of Fig. 4. The MTANNs could not output lower values for the relatively large nodules with spiculation, because these nodules were similar to malignant nodules in their appearance. Therefore, one of the limitations of the MTANN would be the distinction of malignant nodules from the benign nodules which would be similar to malignant nodules in their appearance. This limitation might be the limitation of the LDCT as well.

V. DISCUSSION

We investigated the effect of the change in the number of training nodules on the performance of the MTANN. Seven sets with different numbers of typical malignant and benign nodules were selected from the entire database according to their visual appearance, so that a set of a smaller number of training nodules was a subset of a larger number of training nodules. We trained seven MTANNs with the seven sets with different numbers of nodules from four (two malignant nodules and two benign nodules) to 60 (30 malignant nodules and 30 benign nod-

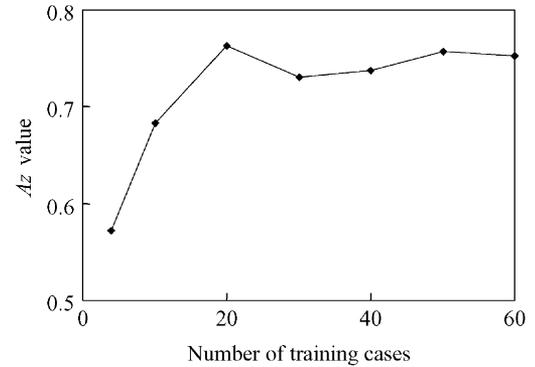


Fig. 12. Effect of the change in the number of training nodules (malignant and benign nodules) on the performance of the MTANN.

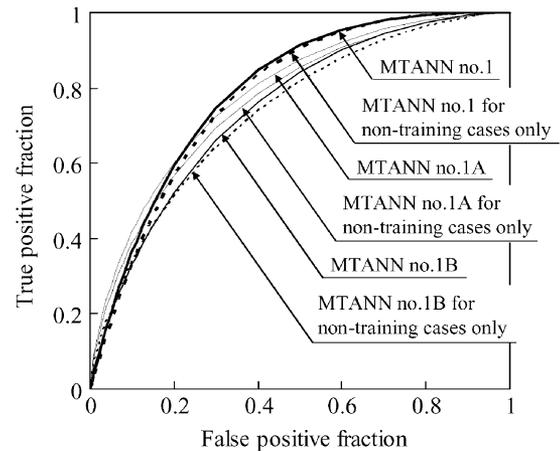


Fig. 13. Effect of the change of a set of training nodules (malignant and benign nodules) on the performance of the MTANN.

ules). The performance of the MTANNs was evaluated by use of ROC analysis. Figure 12 shows the results for nontraining nodules, i.e., the 60 training nodules were excluded from the cases for evaluation. There was little increase in the Az value when the number of training nodules was greater than 20 (ten malignant nodules and ten benign nodules). This is the reason for the use of 20 training nodules for the MTANN. This result was consistent with that in [25].

We investigated the effect of the change of training nodules on the performance of the MTANN. We selected two different sets of ten typical malignant nodules and ten small benign nodules overlapping with vessels by the same way as for MTANN no. 1. We trained two MTANNs (MTANN no. 1A and MTANN no. 1B) by use of the two training sets. The ROC curves of MTANN no. 1 and the two MTANNs are shown in Fig. 13. The Az value for MTANN no. 1, that for MTANN no. 1A, and that for MTANN no. 1B were 0.79, 0.78, and 0.77, respectively. The performance of the MTANNs trained with different cases differed slightly. Less careful selection of training nodules such as random selection, however, would lower the performance. In addition, the performance of each MTANN for nontraining cases was only slightly lower, as shown in Fig. 13. These results were consistent with those in [25], [27].

An ANN generally requires training with a large number of cases, because the ANN has a number of parameters (weights) to be adjusted by the training cases. The inputs of the ANN may

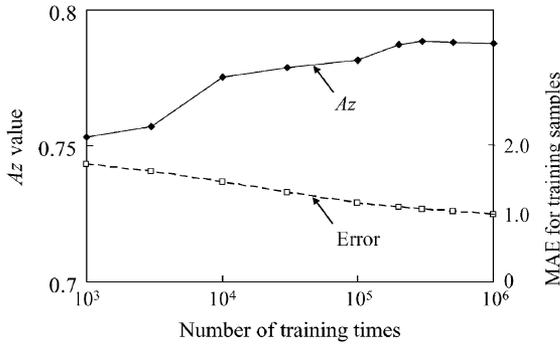


Fig. 14. Learning curve of MTANN no. 1 and the effect of the number of training times on the generalization performance of the MTANN.

often be the image features, which would include, in general, some noise due to the fluctuation in the feature extraction. The MTANN was able to be trained with a small number of training nodules (ten malignant nodules and ten benign nodules). The key to this high generalization ability might be the division of one nodule image into a large number ($361 = 19 \times 19$) of subregions. The 361 subregions could include various parts of the nodule, various nodule margins with different orientations, and also various parts of vessels that overlap with the nodule. This allowed us to train the MTANN not on a case basis, but on a subregion basis. We treated the distinction between malignant and benign nodules as an image-processing task, in other words, as nonlinear filtering that performs both enhancement of malignant nodules and suppression of benign nodules. The MTANN does not see the whole nodule, but rather the image features in the subregions such as the mean CT value, the contrast, the gray-level shape, edges, and texture. The results might suggest that there are some consistent features representing the variability of cancers in the subregions. Thus, massive training with a large number ($7\,220 = 361 \times 20$) of subregions would contribute to the proper determination of the parameters. Moreover, direct use of pixel values instead of image features as the inputs would keep one from mixing the input information with the noise due to the fluctuation in the feature extraction based on segmented nodules. The above would be the reasons for the high generalization ability of the MTANN.

We investigated the property of the MTANN regarding an overtraining issue. Figure 14 shows a learning curve (mean absolute error (MAE) for training samples) of MTANN no. 1 and the effect of the number of training times on the generalization performance (Az values for nontraining cases). There was little increase in Az value when the number of training times was greater than 200 000, and there was a slight decrease at 1 000 000 times. This is the reason for determining the condition for stopping of the training at 500 000. Note that significant overtraining was not seen. This result was consistent with that in [25].

We investigated the effect of parameter change on the performance of the MTANN. The standard deviation σ of the 2-D Gaussian weighting function for scoring for MTANN no. 1 was changed, and the performance for nontraining cases was obtained, as shown in Fig. 15. Because the performance was the highest at a standard deviation of 7.5, we used this value for

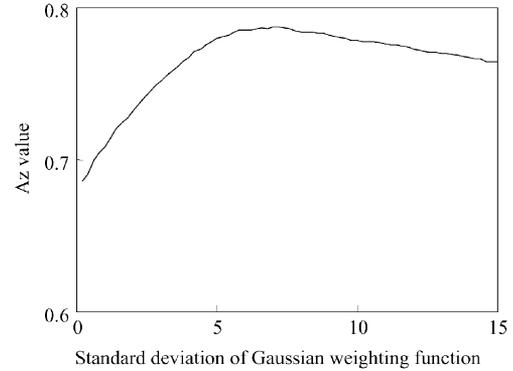


Fig. 15. Effect of the change in the standard deviation σ of the 2-D Gaussian weighting function for scoring on the performance of MTANN no. 1.

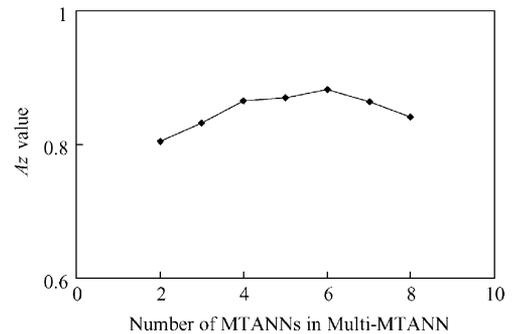


Fig. 16. Effect of the change in the number of MTANNs in the multi-MTANN on the performance of our scheme in the round-robin test.

MTANN no. 1. Thus, the performance was not sensitive to the standard deviation σ . This result was consistent with that in the distinction between nodules and nonnodules in CT images in [25]. Similarly, we determined the standard deviations for other MTANNs to be 7.5 or 8.0.

We investigated the effect of the change in the number of MTANNs in the multi-MTANN on the performance of our scheme. The performance was evaluated by ROC analysis. Note that the number of MTANNs corresponds to the number of input units in the integration ANN. The integration ANN was evaluated by use of a round-robin test. Figure 16 shows the Az values of our schemes with various numbers of MTANNs. A set of a larger number of MTANNs included a set of a smaller number of MTANNs, e.g., two MTANNs were MTANN nos. 1 and 2, and three MTANNs were MTANNs nos. 1, 2, and 3. The seventh group included small nodules with spiculation, and the eighth group included small nodules overlapping with small vessels on a light background. The results show that the performance of our scheme was the highest when the number of MTANNs was six. The differences of the Az value for our scheme consisting of six MTANNs from that for our scheme consisting of two MTANNs, three MTANNs, and eight MTANNs were statistically significant. The use of six MTANNs with this grouping of benign nodules would be a better choice for realization of this CAD scheme, because this result was obtained based on a relatively large database from a lung cancer screening program on 7847 screenees. However, a test on a larger database will produce a much more reliable result for this choice. Because the integration ANN is

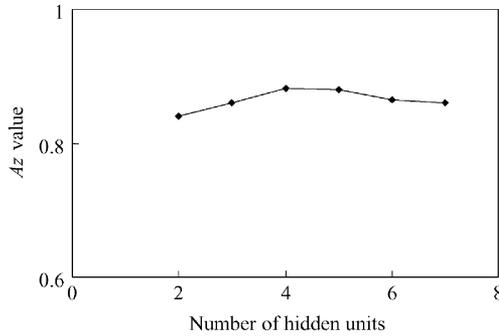


Fig. 17. Effect of the change in the number of hidden units in the integration ANN on the performance of our scheme in the round-robin test.

a conventional multilayer ANN, the integration ANN should have the general properties of a conventional classifier. Scores from the MTANNs can be considered to be the input features for the integration ANN. As the dimensionality of the feature space increases subject to the number of input features, the number of training samples required for a classifier increases exponentially; this is referred to as the curse of dimensionality [40]. Although the increase in the number of input features provides more information for classification, the curse of dimensionality causes a decline of the classification accuracy. Furthermore, an additional input feature may not contribute very much to an improvement in the classification accuracy, because a new input feature can correlate to other features. As a result, the classification accuracy of a classifier first grows and then declines as the number of input features increases, which is referred to as the Hughes phenomenon [41]. Therefore, there exists an optimal number of input units in the integration ANN, in other words, the optimal number of MTANNs.

We also investigated the effect of the change in the number hidden units in the integration ANN in our scheme. The integration ANN was evaluated by use of the round-robin test. The number of MTANNs (i.e., the number of input units) was six. Figure 17 shows the performance of our scheme with various numbers of hidden units. The performance was not very sensitive to the number of hidden units.

We compared the performance of the integration ANN with that of another method for combining the outputs of the multi-MTANN. An average operation is often used for combining multiple classifiers, and would give better results compared to the majority logic [38], [39]. The average operation was performed on the scores from the six MTANNs in the multi-MTANN. The performance of the multi-MTANN combined with the average operation is shown in Fig. 10. The performance of the average operation (Az value of 0.822) was apparently inferior to that of the integration ANN.

We have used the logical AND operation to combine the scores from each MTANN in the multi-MTANN for application to false-positive reduction in CAD for lung nodule detection on LDCT [25], because the scheme should output a binary value, i.e., a true positive (nodule) or a false positive (nonnodule) for the purpose of reduction of false positives. For radiologists' classification task such as distinction between benign and malignant nodules in LDCT, however, we plan to display the computer-estimated likelihood of malignancy

with a proper marker on a nodule rather than only a simple marker indicating a malignant nodule as an aid in radiologists' decision-making. The use of the integration ANN allows us to provide the computer-estimated likelihood of malignancy which is a continuous value, whereas the logical AND operation cannot output a continuous value. The computer-estimated likelihood of malignancy can be calculated from the output values of the integration ANN in our scheme by use of the relationship defined in [42]. The output values of the integration ANN can be transformed to the computer-estimated likelihood of malignancy by use of the maximum-likelihood estimated binormal model in ROC analysis. The computer-estimated likelihood of malignancy was defined [42] as

$$LM(d) = \frac{\lambda M(d)}{\lambda M(d) + (1 - \lambda)B(d)} \quad (4)$$

where d is the latent decision variable, $M(d)$ is the probability density function of d for actually malignant nodules, $B(d)$ is the probability density function of d for actually benign nodules, and λ is the prevalence of malignant nodules in the population studied. In addition, the output of the integration ANN can be employed as a binary decision by use of a threshold value. Thus, our scheme can be used for providing either the computer-estimated likelihood of malignancy of a nodule or a malignant nodule marker by combining our scheme with a detection scheme [25].

For evaluating radiologists' performance in distinction between benign and malignant nodules on LDCT, Li *et al.* have performed an observer study [43], [44]. They randomly selected 20 malignant nodules and 20 benign nodules from the database used in this study. Sixteen radiologists (twelve attending radiologists and four radiology residents) participated in this study. They used ROC analysis for evaluation of the performance of the radiologists. The radiologists were asked whether the nodule was benign or malignant, and then they marked their confidence level regarding the likelihood of malignancy by using a continuous rating scale. An average Az value of 0.72 (0.75 for attending radiologists and 0.62 for residents) was obtained by the 16 radiologists in the observer study, whereas our scheme achieved a higher Az value (0.882) than did the radiologists. Therefore, we expect that our scheme would be useful in improving radiologists' classification accuracy.

Researchers have developed computerized schemes for distinction between benign and malignant lesions in chest radiographs [45], [46], mammograms [47]–[51], and CT images [52]–[54], and also computerized schemes for detection of nodules in CT [55], [56]. Aoyama *et al.* have developed a computerized scheme for distinction between benign and malignant lung nodules in LDCT. Table II shows the difference between Aoyama's scheme and our scheme based on the MTANN. The performance (Az value of 0.882) of our scheme was greater than that of Aoyama's scheme of 0.828 [54] for the same cases in the same database by a statistical significant level (two-tailed p value = 0.0068) [57]. The 95% confidence intervals of the Az value of our scheme were 0.834 and 0.908, which were greater than the Az value of Aoyama's scheme. Aoyama's scheme was based on the following four steps: determination of the locations of nodules, segmentation of the nodules, feature

TABLE II
DIFFERENCE BETWEEN AOYAMA'S SCHEME AND OUR SCHEME
BASED ON THE MTANN

	Aoyama's segmentation-based scheme	MTANN-based scheme
Segmentation	Radial search of edge candidates based on edge magnitude and contour smoothness	No segmentation
Feature analysis	Three gray-level-based features, two edge-based features, and one morphological feature, plus clinical information	Multi-MTANN (pixel-based determination of likelihood of malignancy from sub-regions)
Classification	Linear discriminant analysis	Integration ANN
Performance	0.828	0.882

extraction and analysis of the segmented nodules, and linear discriminant analysis [58] for distinction between benign and malignant nodules. The first step was performed manually by a chest radiologist. The second step of segmentation was performed by use of the radial search of edge candidates based on edge magnitude and contour smoothness for determining the regions of the nodules. Note that we made a particular use of the technical term, segmentation, as a technique for determining the region of a target object in images by following the definition in the field of image processing [59]. The accuracy of the segmentation can affect the accuracy of the feature extraction and analysis, and therefore, the final accuracy of classification. The features of a nodule included three gray-level-based features, two edge-based features, a morphological feature, and clinical information. We believe that accurate segmentation is difficult; therefore, incorrect segmentation can occur for complicated patterns such as nodules overlapping with vessels and subtle opacities like GGO. However, the use of MTANN does not require the step of the segmentation, but only image data and nodule locations directly. Therefore, there is no room for errors due to incorrect segmentation when the MTANN is employed. We believe that this is a major advantage of the MTANN for classification of lung nodules in CT.

Our method based on a multi-MTANN with the integration ANN can be extended to the use of multiple sections. By applying our trained scheme section by section, output values for a nodule in multiple sections can be obtained. By combining the output values, the score for the nodule can be obtained. The score can be determined by use of the following possible four combining methods: 1) the maximum value among the output values in multiple sections, 2) the minimum value among the output values in multiple sections, 3) an average value of the output values in multiple sections, and 4) a weighted average value of the output values in multiple sections by use of the effective diameter of the nodule at each section as the weighting factor. We expect that the performance would be improved by incorporating of the information in multiple sections.

CT images with different section thickness have been used for lung cancer screening in different medical institutions [2]–[13]; a protocol of section thickness adequate for lung cancer screening has not yet been determined. Many institutions used CT images with a 10 mm section thickness for lung cancer screening [3]–[6], [8], [11]–[13], and some other institutions used a 5 mm section thickness [7], [9]. Some institutions used 10 mm and 2.5 mm [10], but the section thickness is still a controversial issue. We believe

that, if our scheme is applied to CT images with thinner sections, e.g., 5 mm or 2.5 mm, the performance can be improved, because these CT images contain more information on both malignant and benign nodules. Training of the MTANNs with thin-section CT images may be required for accurate classification. Our scheme can be applied to thin-section CT images section by section. With thin-section CT images acquired with a multidetector CT system, smaller benign nodules would be found compared to those in thick-section CT images. Because our scheme was effective for small benign nodules, as shown in Fig. 6, our scheme would be effective for a database of thin-section CT images.

The nodules do not need to be positioned exactly at the center of the ROI. The distributions in the output images are generally broad, as shown in Fig. 7. The broad output distribution and the scoring method with a relatively broad Gaussian weighting function allowed the MTANN to be robust against the change of the location of the nodule. For example, the centers of the nodules in the third ROI in the pure GGO row, the fourth ROI in the mixed GGO row, and the second ROI in the solid nodule row are fairly far from the centers of the ROIs. These nodules were distinguished well from benign nodules by use of our scheme.

We considered the use of our scheme in the case where a radiologist determines the nodule locations. We expect that a radiologist is likely to select the section with the greatest nodule size or a similar size rather than a section with small nodule size, because the nodule in the section with small size appears to be small and low-contrast. Therefore, the selection of the section with the greatest nodule size by a radiologist as used in this study would be reasonable for this particular usage. If the section with the second greatest nodule size is selected, the performance of our scheme would not change much, because the distribution in the output image of the MTANN is relatively broad for small nodules, e.g., a small low-contrast nodule in the output image was appropriately represented by light distributions, as shown in the fourth nodule in the pure GGO row in Fig. 7.

Training of the multi-MTANN took a long time, i.e., about 30 h for each MTANN. There are many methods [60]–[64] for accelerating the convergence speed of the BP algorithm. These methods include learning rate adaptation [60]–[62], training using a Hessian matrix of the cost function [63], and learning-rate optimization [64]. These methods can be applied to our modified BP algorithm, and the time for training can be shortened by use of these methods. Training with our modified BP algorithm can be trapped at local minima, because our modified BP algorithm was based on the BP algorithm. There are many methods [65]–[68] for avoiding local minima for the BP algorithm. By use of these methods, the performance of the MTANN might be improved by avoiding possible local minima.

We considered the generalizability of our scheme and the results in this study. There are two major factors which can affect the generalizability of the results in evaluation of a computerized scheme [69]–[72]: 1) the quality and quantity of the database used and 2) the testing methodology used for evaluating the scheme. Because our database used in this study was relatively large, containing 76 confirmed primary cancers, obtained from a lung cancer screening program on 7 847 screenees for three years. Therefore, we believe that the quality and quantity of our database were appropriate. However, evaluation (and training)

with a larger database will produce more reliable results. Regarding testing methodology, we performed a hold-out testing method on the MTANNs, i.e., we excluded all training cases for the MTANNs from the evaluation. Because the number of training cases for the MTANN was very small, the selection of training cases could affect the performance of the MTANN. The variation in the performances of the MTANNs trained with different training cases was small, as shown in Fig. 13. Therefore, this would not be a problem for generalizability. Because we determined some parameters such as the standard deviation of the 2-D Gaussian weighting function for scoring, the number of MTANNs in the multi-MTANN, and the number of hidden units in the integration ANN by use of the entire database, some biases can be included in the performance [71]. However, the differences in the performance with respect to these parameters were not large, as shown in Figs. 15–17. When the standard deviation, the number of MTANNs, and the number of hidden units differed by 20%, one (approximately 20%), and one (approximately 20%), respectively, the Az value changed by 0.4%, 1.7%, and 1.3% on average, respectively. Thus, the performance was not sensitive to these parameters. Therefore, the biases which may be included during the design process should be small. A possible weakness would be a round-robin test (also referred to as a jackknife test) used for evaluating the integration ANN, because a round-robin test is not an independent test, but a test with resampling. However, many researchers have accepted and used a round-robin test in their studies [45]–[48], [50]–[54]. In addition, according to Fukunaga *et al.*'s findings [69], a round-robin test would produce a pessimistically biased performance compared to the “true” performance (i.e., the performance of a classifier designed with the true population and tested with the true population). Therefore, we trust that our results are reliable, and we expect that results similar to those presented in this paper can be obtained when our scheme is applied to different databases.

We have experienced that ANN models which use a large number of subregions can be trained with a very limited number of cases, including neural filters, neural edge enhancers, and MTANNs. The neural filter and the neural edge enhancer were trained with six images (angiograms) [15] or one image (gastrointestinal radiograph) [23], and one image (landscape) [17] or three images (ventriculograms) [18], respectively. The MTANN was able to be trained with 20 cases (CT images containing ten nodules and ten nonnodules) [24] or 12 cases (CT images containing six nodules and six nonnodules) [25]. A convolution ANN [73] is a different type of ANN, but operates on image data directly. A convolution ANN was trained with 28 cases (including nodule cases and normal cases) for differentiation between nodules and nonnodules in chest radiographs. Although the number of training cases was 28 for each of the training subsets in the cross validation scheme, the number of nodules was around 25 and the number of false positives was around 75 in each training subset. The common features of the above ANN models, including MTANNs and a convolution ANN, are the direct use of image data and the use of a large number of subregions (or subimages) extracted from cases (images) for training.

In order to gain insight into the training of the MTANN, we analyzed the information used by the MTANN. The input of the

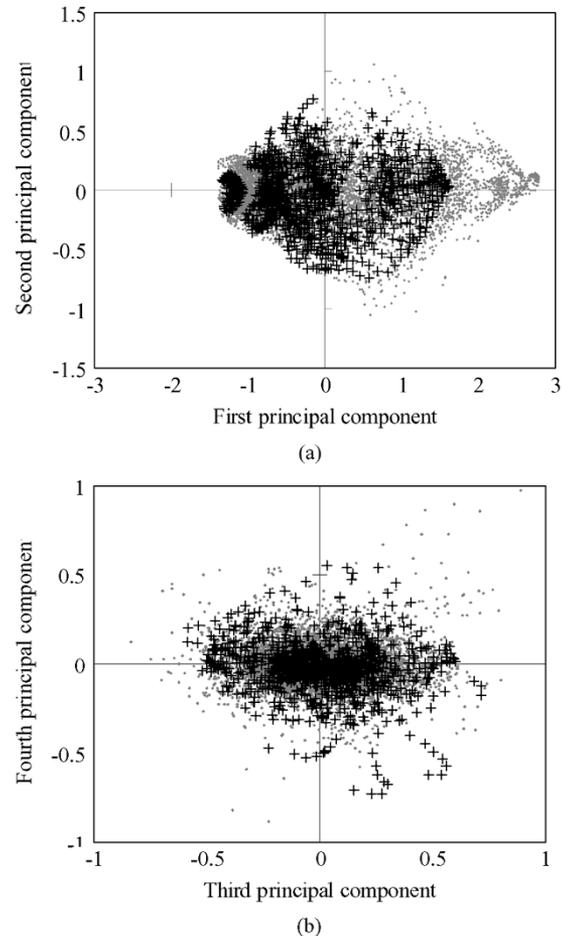


Fig. 18. Distributions of samples extracted from the 10 training malignant nodules and all 76 malignant nodules in the database in the principal component (PC) vector space. Black crosses represent samples (subregions) extracted from the training cases. Gray dots represent samples extracted from all cases in the database. (a) Relationship between the first and second PCs. (b) Relationship between the third and fourth PCs.

MTANN can be considered as an 81-dimensional (81-D) input vector. In the MTANN approach, each case (nodule image) is divided into a large number (361) of subregions. Each subregion corresponds to the 81-D input vector. If a large number of 81-D input vectors obtained from the training cases (e.g., ten malignant nodules) approximate those obtained from all cases in the database (i.e., 76 malignant nodules), the MTANN trained with these training cases can potentially have a high generalization ability. Because it is difficult to visualize and compare all 81 dimensions of the input vector, we employed principal-component analysis (PCA, also referred to as Karhune-Loeve analysis) [74] for reducing the dimensions. We applied PCA to 81-D vectors obtained from all 76 malignant nodules. Figure 18(a) and (b) shows the distributions of samples (subregions) extracted from the ten training malignant nodules and all 76 malignant nodules in the database in the principal component (PC) vector space. Only the first to fourth PCs are shown in the figures, because the cumulative contribution rate of the fourth PC is 0.974, i.e., the figures represent 97.4% of all data. The result showed that the ten training cases represent the 76 cases fairly well except for the right portion of the distribution in the relationship between the first and second PCs in figure (a). The right portion of the distribution is very sparse, containing only 6% of all samples.

This does not mean that the training nodules do not cover 6% of the 76 nodules, but that the training nodules do not cover, on average, 6% of the components of each nodule. Because all components of each nodule are combined with the scoring method in the MTANN, the noncovered 6% of components would not be critical at all for the classification accuracy. Thus, the division of each nodule case into a large number of subregions enriched the variations in the feature components of nodules, and therefore contributed to the generalization ability of the MTANN.

VI. CONCLUSION

Our computerized scheme for distinction between benign and malignant nodules based on the multi-MTANN incorporated with the integration ANN achieved a relatively high Az value of 0.882, and would be useful in assisting radiologists in the diagnosis of lung nodules in LDCT by reducing the number of “unnecessary” HRCTs and/or biopsies.

ACKNOWLEDGMENT

The authors are grateful to Q. Li, Ph.D., and M. Aoyama, Ph.D., for preparing images, H. Abe, MD, Ph.D., for his clinical advice, J. Shiraishi, Ph.D., H. Arimura, Ph.D., H. Takizawa, Ph.D., H. MacMahon, MD, R. Engelmann, MS, C. Muramatsu, BA, S. G. Armato, Ph.D., and M. L. Giger, Ph.D., for their valuable suggestions, to C. E. Metz, Ph.D., for the use of the LABROC5 program, and E. F. Lanzl for improving the manuscript. They also thank the Associate Editor and anonymous referees for their suggestions. K. Doi is a shareholder in R2 Technology, Inc., Sunnyvale, CA.

REFERENCES

- [1] A. Jemal, T. Murray, A. Samuels, A. Ghafoor, E. Ward, and M. J. Thun, “Cancer statistics, 2003,” *CA Cancer J. Clin.*, vol. 53, no. 1, pp. 5–26, Jan. 2003.
- [2] O. S. Miettinen and C. I. Henschke, “CT screening for lung cancer: Coping with nihilistic recommendations,” *Radiol.*, vol. 221, no. 3, pp. 592–596, Dec. 2001.
- [3] M. Kaneko, K. Eguchi, H. Ohmatsu, R. Kakinuma, T. Naruke, K. Suenasu, and N. Moriyama, “Peripheral lung cancer: Screening and detection with low-dose spiral CT versus radiography,” *Radiol.*, vol. 201, no. 3, pp. 798–802, Dec. 1996.
- [4] S. Sone, S. Takashima, F. Li, Z. Yang, T. Honda, Y. Maruyama, M. Hasegawa, T. Yamada, K. Kubo, K. Hanamura, and K. Asakura, “Mass screening for lung cancer with mobile spiral computed tomography scanner,” *Lancet*, vol. 351, pp. 1242–1245, Apr. 1998.
- [5] C. I. Henschke, D. I. McCauley, D. F. Yankelevitz, D. P. Naidich, G. McGuinness, O. S. Miettinen, D. M. Libby, M. W. Pasmantier, J. Koizumi, N. K. Altorki, and J. P. Smith, “Early lung cancer action project: Overall design and findings from baseline screening,” *Lancet*, vol. 354, pp. 99–105, Jul. 1999.
- [6] C. I. Henschke *et al.*, “Early lung cancer action project: Initial finding on repeat screening,” *Cancer*, vol. 92, no. 1, pp. 153–159, July 2001.
- [7] S. J. Swensen, J. R. Jett, T. E. Hartman, D. E. Midthun, J. A. Sloan, A. M. Sykes, G. L. Aughenbaugh, and M. A. Clemens, “Lung cancer screening with CT: Mayo Clinic experience,” *Radiol.*, vol. 226, no. 3, pp. 756–761, Mar. 2003.
- [8] H. Rusinek, D. P. Naidich, G. McGuinness, B. S. Leitman, D. I. McCauley, G. A. Krinsky, K. Clayton, and H. Cohen, “Pulmonary nodule detection: Low-dose versus conventional CT,” *Radiol.*, vol. 209, no. 1, pp. 243–249, Oct. 1998.
- [9] K. Garg, R. L. Keith, T. Byers, K. Kelly, A. L. Kerzner, D. A. Lynch, and Y. E. Miller, “Randomized controlled trial with low-dose spiral CT for lung cancer screening: Feasibility study and preliminary results,” *Radiol.*, vol. 225, no. 2, pp. 506–510, Nov. 2002.
- [10] C. I. Henschke, D. F. Yankelevitz, D. P. Naidich, D. I. McCauley, G. McGuinness, D. M. Libby, J. P. Smith, M. W. Pasmantier, and O. S. Miettinen, “CT screening for lung cancer: Suspiciousness of nodules according to size on baseline scans,” *Radiol.*, vol. 231, no. 1, pp. 164–168, Apr. 2004.
- [11] S. Sone, F. Li, Z. G. Yang, T. Honda, Y. Maruyama, S. Takashima, M. Hasegawa, S. Kawakami, K. Kubo, M. Haniuda, and T. Yamada, “Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner,” *Br. J. Cancer*, vol. 84, no. 1, pp. 25–32, Jan. 2001.
- [12] T. Nawa, T. Nakagawa, S. Kusano, Y. Kawasaki, Y. Sugawara, and H. Nakata, “Lung cancer screening using low-dose spiral CT,” *Chest*, vol. 122, no. 1, pp. 15–20, July 2002.
- [13] F. Li, S. Sone, H. Abe, H. MacMahon, S. G. Armato, and K. Doi, “Lung cancer missed at low-dose helical CT screening in a general population: Comparison of clinical, histopathologic, and imaging findings,” *Radiol.*, vol. 225, no. 3, pp. 673–683, Dec. 2002.
- [14] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, “Noise reduction of medical X-ray image sequences using a neural filter with spatiotemporal inputs,” in *Proc. Int. Symp. Noise Reduction for Imag. and Comm. Systems*, Nov. 1998, pp. 85–90.
- [15] —, “Neural filter with selection of input features and its application to image quality improvement of medical image sequences,” *IEICE Trans. Inform. Syst.*, vol. E85-D, no. 10, pp. 1710–1718, Oct. 2002.
- [16] K. Suzuki, I. Horiba, and N. Sugie, “Neural edge detector—A good mimic of conventional one yet robust against noise,” in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, Jun. 2001, vol. 2085, pp. 303–310.
- [17] —, “Neural edge enhancer for supervised edge enhancement from noisy images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1582–1596, Dec. 2003.
- [18] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, “Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector,” *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 330–339, Mar. 2004.
- [19] K. Suzuki, I. Horiba, and N. Sugie, “Training under achievement quotient criterion,” in *Neural Networks for Signal Processing X*. Piscataway, NJ: IEEE Press, 2000, pp. 537–546.
- [20] —, “Simple unit-pruning with gain-changing training,” in *Neural Networks for Signal Processing XI*. Piscataway, NJ: IEEE Press, 2001, pp. 153–162.
- [21] —, “Designing the optimal structure of a neural filter,” in *Neural Networks for Signal Processing VIII*. Piscataway, NJ: IEEE Press, 1998, pp. 323–332.
- [22] —, “A simple neural network pruning algorithm with application to filter synthesis,” *Neural Process. Lett.*, vol. 13, no. 1, pp. 43–53, Feb. 2001.
- [23] K. Suzuki, “Determining the receptive field of a neural filter,” *J. Neural Eng.*, vol. 1, no. 4, pp. 228–237, Dec. 2004.
- [24] K. Suzuki, I. Horiba, and N. Sugie, “Efficient approximation of neural filters for removing quantum noise from images,” *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1787–1799, Jul. 2002.
- [25] K. Suzuki, S. G. Armato, F. Li, S. Sone, and K. Doi, “Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose CT,” *Med. Phys.*, vol. 30, no. 7, pp. 1602–1617, Jul. 2003.
- [26] —, “Effect of a small number of training cases on the performance of massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose CT,” in *Proc. SPIE (Medical Imaging)*, vol. 5032, San Diego, CA, May 2003, pp. 1355–1366.
- [27] H. Arimura, S. Katsuragawa, K. Suzuki, F. Li, J. Shiraishi, S. Sone, and K. Doi, “Computerized scheme for automated detection of lung nodules in low-dose CT images for lung cancer screening,” *Acad. Radiol.*, vol. 11, no. 6, pp. 617–629, Jun. 2004.
- [28] K. Suzuki and K. Doi, “Characteristics of a massive training artificial neural network (MTANN) in the distinction between lung nodules and vessels in CT images,” in *Computer Assisted Radiology and Surgery (CARS)*, Chicago, IL, Jun. 2004, pp. 923–928.
- [29] K. Suzuki, J. Shiraishi, H. Abe, H. MacMahon, and K. Doi, “False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network,” *Acad. Radiol.*, vol. 12, no. 2, pp. 191–201, Feb. 2005.
- [30] K. Suzuki, I. Horiba, K. Ikegaya, and M. Nanki, “Recognition of coronary arterial stenosis using neural network on DSA system,” *Syst. Comput. Jpn.*, vol. 26, no. 8, pp. 66–74, Aug. 1995.
- [31] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations of back-propagation errors,” *Nature*, vol. 323, pp. 533–536, 1986.
- [32] —, “Learning internal representations by error propagation,” in *Parallel Distributed Processing*. Cambridge, MA: MIT Press, 1986, vol. 1, pp. 318–362.
- [33] K. Funahashi, “On the approximate realization of continuous mappings by neural networks,” *Neural Netw.*, vol. 2, pp. 183–192, 1989.

- [34] A. R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 930–945, May 1993.
- [35] C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.*, vol. 21, pp. 720–733, 1986.
- [36] 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, no. 9, pp. 1033–1053, May 1998.
- [37] J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiol.*, vol. 148, no. 3, pp. 839–843, Sep. 1983.
- [38] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
- [39] J. Kittler and F. M. Alkoot, "Sum versus vote fusion in multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 1, pp. 110–115, Jan. 2003.
- [40] R. Bellman, *Adaptive Control Processes, A Guided Tour*. Princeton, NJ: Princeton Univ. Press, 1961.
- [41] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, pp. 55–63, 1968.
- [42] Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, "Improving breast cancer diagnosis with computer-aided diagnosis," *Acad. Radiol.*, vol. 6, no. 1, pp. 22–33, Jan. 1999.
- [43] Q. Li, M. Aoyama, F. Li, S. Sone, H. MacMahon, and K. Doi, "Potential clinical usefulness of an intelligent computer-aided diagnostic scheme for distinction between benign and malignant pulmonary nodules in low-dose CT scans," *Radiol.*, vol. 225(P), no. 2, pp. 534–535, Nov. 2002.
- [44] Q. Li, F. Li, S. Katsuragawa, J. Shiraiishi, H. MacMahon, S. Sone, and K. Doi, "Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules," *Med. Phys.*, vol. 30, no. 10, pp. 2584–2593, Oct. 2003.
- [45] K. Nakamura, H. Yoshida, R. Engelmann, H. MacMahon, S. Katsuragawa, T. Ishida, K. Ashizawa, and K. Doi, "Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules by use of artificial neural networks," *Radiol.*, vol. 214, no. 3, pp. 823–830, Mar. 2000.
- [46] M. Aoyama, Q. Li, S. Katsuragawa, H. MacMahon, and K. Doi, "Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images," *Med. Phys.*, vol. 29, no. 5, pp. 701–708, May 2002.
- [47] Y. Jiang, R. M. Nishikawa, D. E. Wolverton, C. E. Metz, M. L. Giger, R. A. Schmidt, C. J. Vyborny, and K. Doi, "Malignant and benign clustered microcalcifications: Automated feature analysis and classification," *Radiol.*, vol. 198, no. 3, pp. 671–678, Mar. 1996.
- [48] H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, K. L. Lam, D. D. Adler, and M. M. Goodsitt, "Computerized classification of malignant and benign microcalcifications on mammograms: Texture analysis using an artificial neural network," *Phys. Med. Biol.*, vol. 42, no. 3, pp. 549–567, Mar. 1997.
- [49] Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign mass lesions on digitized mammograms," *Acad. Radiol.*, vol. 5, pp. 155–168, 1998.
- [50] L. Hadjiiski, B. Sahiner, H.-P. Chan, N. Petrick, and M. Helvie, "Classification of malignant and benign masses based on hybrid ART2LDA approach," *IEEE Trans. Med. Imag.*, vol. 8, no. 12, pp. 1178–1187, Dec. 1999.
- [51] B. Sahiner, N. Petrick, H. P. Chan, L. M. Hadjiiski, C. Paramagul, M. A. Helvie, and M. N. Gurcan, "Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1275–1284, Dec. 2001.
- [52] Y. Matsuki, K. Nakamura, H. Watanabe, T. Aoki, H. Nakata, S. Katsuragawa, and K. Doi, "Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution CT: Evaluation with receiver operating characteristic analysis," *AJR*, vol. 178, pp. 657–663, Mar. 2002.
- [53] M. F. McNitt-Gray, E. M. Hart, N. Wyckoff, J. W. Sayre, J. G. Goldin, and D. R. Aberle, "A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results," *Med. Phys.*, vol. 26, no. 6, pp. 880–888, Jun. 1999.
- [54] M. Aoyama, Q. Li, S. Katsuragawa, F. Li, S. Sone, and K. Doi, "Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images," *Med. Phys.*, vol. 30, no. 3, pp. 387–394, Mar. 2003.
- [55] S. Chang, H. Emoto, D. N. Metaxas, and L. Axel, "Pulmonary micronodule detection from 3D chest CT," *Med. Image Comput. Comput.-Assist. Intervention*, pt. I, pp. 821–828, Sep. 2004.
- [56] A. Farag, A. El-Baz, G. G. Gimel'farb, R. Falk, and S. G. Hushek, "Automatic detection and recognition of lung abnormalities in helical CT images using deformable templates," *Med. Image Comput. Comput.-Assist. Intervention*, pt. I, pp. 856–864, Sep. 2004.
- [57] C. E. Metz, B. A. Herman, and C. A. Roe, "Statistical comparison of two ROC curve estimates obtained from partially-paired datasets," *Med. Decision Making*, vol. 18, pp. 110–121, 1998.
- [58] P. A. Lachenbruch, *Discriminant Analysis*. New York: Hafner, 1975, pp. 1–39.
- [59] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, 2nd ed. San Diego, CA: Academic, 1982, vol. 2, pp. 55–190.
- [60] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Netw.*, vol. 1, pp. 295–307, 1988.
- [61] T. P. Vogl, J. K. Mangis, A. K. Rigler, W. T. Zink, and D. L. Alkon, "Accelerating the convergence of the back-propagation method," *Biol. Cybern.*, vol. 59, pp. 257–263, 1988.
- [62] R. Battiti, "Accelerated backpropagation learning: Two optimization methods," *Complex Syst.*, vol. 3, pp. 331–342, 1989.
- [63] Y. LeCun, I. Kanter, and S. A. Solla, "Second order properties of error surfaces: Learning time and generalization," in *Advances in Neural Information Processing Systems*, vol. 3. Cambridge, MA, 1991, pp. 918–924.
- [64] Y. LeCun, P. Y. Simaard, and B. Pearlmutter, "Automatic learning rate maximization by on-line estimation of the Hessian's eigenvectors," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1993, vol. 5, pp. 156–163.
- [65] Y. Shang and B. W. Wah, "Global optimization for neural network training," *Computer*, vol. 29, no. 3, pp. 45–54, 1996.
- [66] S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture," in *Advances in Neural Information Processing Systems 2*. San Mateo, CA: Morgan Kaufmann, 1990, pp. 524–532.
- [67] W. Finnoff, "Diffusion approximations for the constant learning rate backpropagation algorithm and resistance to local minima," *Neural Comput.*, vol. 6, no. 2, pp. 285–295, 1994.
- [68] P. Baldi and K. Hornik, "Neural networks and principal component analysis: Learning from examples without local minima," *Neural Netw.*, vol. 2, no. 1, pp. 53–58, 1989.
- [69] K. Fukunaga and R. R. Hayes, "Effects of sample size on classifier design," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 11, no. 10, pp. 873–885, 1989.
- [70] H. P. Chan, B. Sahiner, R. F. Wagner, and N. Petrick, "Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers," *Med. Phys.*, vol. 26, no. 12, pp. 2654–2668, Dec. 1999.
- [71] B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner, and L. Hadjiiski, "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size," *Med. Phys.*, vol. 27, no. 7, pp. 1509–1522, Jul. 2000.
- [72] H. P. Chan, B. Sahiner, and L. Hadjiiski, "Sample size and validation issues on the development of CAD systems," in *Proc. Computer Assisted Radiology and Surgery (CARS)*, Chicago, IL, June 2004, pp. 872–877.
- [73] S. B. Lo, H. P. Chan, J. S. Lin, H. Li, M. Freedman, and S. K. Mun, "Artificial convolution neural network for medical image pattern recognition," *Neural Netw.*, vol. 8, no. 7/8, pp. 1201–1214, 1995.
- [74] E. Oja, *Subspace Methods of Pattern Recognition*. Letchworth, U.K.: Research Studies, 1983.