Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography

Kenji Suzuki,[a] Samuel G. Armato III, Feng Li, Shusuke Sone,[b] and Kunio Doi
Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, The University of Chicago, Chicago, Illinois 60637

(Received 23 May 2002; accepted for publication 10 April 2003; published 20 June 2003)

In this study, we investigated a pattern-recognition technique based on an artificial neural network (ANN), which is called a massive training artificial neural network (MTANN), for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography (CT) images. The MTANN consists of a modified multilayer ANN, which is capable of operating on image data directly. The MTANN is trained by use of a large number of subregions extracted from input images together with the teacher images containing the distribution for the ''likelihood of being a nodule.'' The output image is obtained by scanning an input image with the MTANN. The distinction between a nodule and a non-nodule is made by use of a score which is defined from the output image of the trained MTANN. In order to eliminate various types of non-nodules, we extended the capability of a single MTANN, and developed a multiple MTANN (Multi-MTANN). The Multi-MTANN consists of plural MTANNs that are arranged in parallel. Each MTANN is trained by using the same nodules, but with a different type of non-nodule. Each MTANN acts as an expert for a specific type of non-nodule, e.g., five different MTANNs were trained to distinguish nodules from various-sized vessels; four other MTANNs were applied to eliminate some other opacities. The outputs of the MTANNs were combined by using the logical AND operation such that each of the trained MTANNs eliminated none of the nodules, but removed the specific type of non-nodule with which the MTANN was trained, and thus removed various types of non-nodules. The Multi-MTANN consisting of nine MTANNs was trained with 10 typical nodules and 10 non-nodules representing each of nine different non-nodule types (90 training non-nodules overall) in a training set. The trained Multi-MTANN was applied to the reduction of false positives reported by our current computerized scheme for lung nodule detection based on a database of 63 low-dose CT scans (1765 sections), which contained 71 confirmed nodules including 66 biopsy-confirmed primary cancers, from a lung cancer screening program. The Multi-MTANN was applied to 58 true positives (nodules from 54 patients) and 1726 false positives (non-nodules) reported by our current scheme in a validation test; these were different from the training set. The results indicated that 83% (1424/1726) of non-nodules were removed with a reduction of one true positive (nodule), i.e., a classification sensitivity of 98.3% (57 of 58 nodules). By using the Multi-MTANN, the false-positive rate of our current scheme was improved from 0.98 to 0.18 false positives per section (from 27.4 to 4.8 per patient) at an overall sensitivity of 80.3% (57/71). [DOI: 10.1118/1.1580485]

Key words: computer-aided diagnosis (CAD), artificial neural network, pattern recognition, lung nodule, false positive

I. INTRODUCTION

Lung cancer continues to rank as the leading cause of cancer death among Americans and is expected to cause 154 900 deaths in the United States in 2002.[1] Some evidence suggests that early detection of lung cancer may allow more timely therapeutic intervention and thus a more favorable prognosis for the patient.[2,3] Accordingly, lung cancer screening programs are being conducted in the United States and Japan[4,5] with low-dose helical computed tomography (CT) as the screening modality. Helical CT, however, generates a large number of images that must be read by radiologists. This may lead to ''information overload'' for the radiologists. Furthermore, radiologists may miss some cancers during interpretation of CT images.[6,7] Therefore, a computer-aided diagnostic (CAD) scheme for detection of lung nodules in low-dose CT images has been investigated as a useful tool for lung cancer screening.

Many investigators have developed a number of methods for the automated detection of lung nodules in CT scans, based on morphological filtering,[8,9] geometric modeling,[10] fuzzy clustering,[11] and gray-level thresholding.[12–17] Giger *et al.*,[12] for example, developed an automated detection scheme based on multiple gray-level thresholding and geometric feature analysis. Armato *et al.*[13–16] extended the method to include a three-dimensional approach combined with linear discriminant analysis.

A major problem with our current CAD scheme for lung nodule detection is a relatively large number of false positives, which could cause difficulty in the clinical application of the CAD scheme. A large number of false positives is likely to confound the radiologist's task of image interpretation, thus lowering the radiologist's efficiency with the CAD scheme. In addition, radiologists may lose their confidence in CAD as a useful tool. Therefore, it is important to reduce the number of false positives as much as possible, while maintaining a high sensitivity.

Our purpose in this study is to develop a pattern-recognition technique based on an artificial neural network (ANN), which is called a massive training artificial neural network (MTANN), for reduction of false positives in computerized detection of lung nodules in low-dose CT images.

II. MATERIALS

A. Database of low-dose CT images

The database used in this study consisted of 101 noninfused, low-dose thoracic helical CT (LDCT) scans acquired from 71 different patients who participated voluntarily in a lung cancer screening program between 1996 and 1999 in Nagano, Japan.[3,18,7] 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, 25 mA (54 scans) or 50 mA (47 scans), 10-mm collimation, and a 10-mm reconstruction interval at a helical pitch of two.[18] The pixel size was 0.586 mm for 83 scans and 0.684 mm for 18 scans. Each reconstructed CT section had an image matrix size of 512×512 pixels. We used 38 of 101 LDCT scans which were acquired from 31 patients as a training set for our CAD scheme. The 38 scans consisted of 1057 sections and contained 50 nodules, including 38 ''missed'' nodules that represented biopsy-confirmed lung cancers and were not reported or misreported during the initial clinical interpretation.[7] The remaining 12 nodules in the scans were classified as ''confirmed benign'' ($n=8$), ''suspected benign'' ($n=3$), or ''suspected malignant'' ($n=1$). The confirmed benign nodules were determined by biopsy or by follow-up over a period of 2 years. The suspected benign nodules were determined by follow-up less than 2 years. The suspected malignant nodule was determined on the basis of results of follow-up diagnostic CT studies; no biopsy results were available. We used 63 of 101 LDCT scans which were acquired from 63 patients as a test set. The 63 scans consisted of 1765 sections and contained 71 nodules, including 66 primary cancers that were determined by biopsy and five confirmed benign nodules that were determined by biopsy or by follow-up over a period of 2 years. The scans included 23 scans from the same 23 patients as those in the training set, which were acquired at a different time (the interval was about 1 year or 2 years). Thus, the training set consisted of 38 LDCT scans including 50 nodules, and the test set consisted of 63 LDCT scans including 71 confirmed nodules.

Figure 1 shows the distributions of nodule sizes for the training set and the test set in our database. The nodule size was determined by an experienced chest radiologist (F.L.), and ranged from 4 mm to 27 mm. When the nodule was present in more than one section, the greatest size was used as the nodule size. Note that the nodules were present in a maximum of three sections. The mean diameter of the 50 nodules in the training set was 12.7 ± 6.1 mm, and that of the 71 nodules in the test set was 13.5 ± 4.7 mm. In the training set, 38% of nodules were attached to the pleura, 22% of nodules were attached to vessels, and 10% of nodules were in the hilum. As to the test set, 30% of nodules were attached to the pleura, 34% of nodules were attached to vessels, and 7% of nodules were in the hilum. Three radiologists (F.L. and two other experienced chest radiologists) determined the nodules in the training set as three categories such as pure ground-glass opacity (pure GGO; 40% of nodules), mixed GGO (28%), and solid nodule (32%); the nodules in the test set were determined as pure GGO (24%), mixed GGO (30%), and solid nodule (46%).

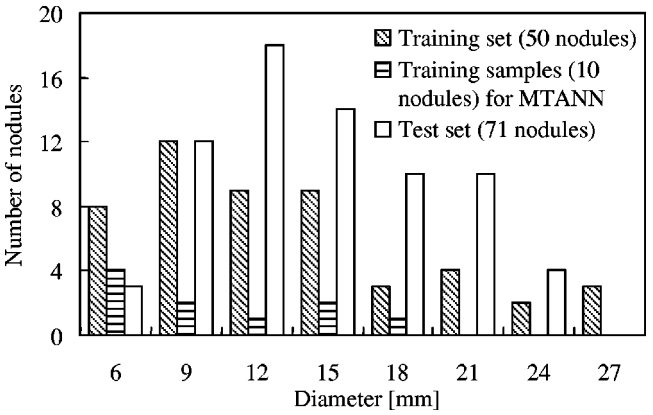


FIG. 1. Distributions of nodule sizes for our database. The training set contained 50 nodules including 38 ''missed'' cancers, and the test set contained 71 confirmed nodules including 66 biopsy-confirmed primary cancers. Ten of the 50 nodules were used for training the MTANN.

B. Current scheme for lung nodule detection in low-dose CT

Technical details of our current scheme have been published previously.[13–16] To summarize the methodology, lung nodule identification proceeds in three phases: two-dimensional (2D) processing, followed by three-dimensional (3D) analysis, and then the application of classifiers. A gray-level-thresholding technique is applied to a 2D section of a CT scan for automated lung segmentation. A multiple gray-level-thresholding technique is applied to the segmented lung volume. Individual structures are identified by grouping of spatially contiguous pixels that remain in the volume at each of 36 gray-level thresholds. A structure is identified as a nodule candidate if the volume of the structure is less than that of a 3-cm-diameter sphere. The categorization of nodule candidates as ''nodule'' or ''non-nodule'' is based on a combination of a rule-based classifier and a series of two linear discriminant classifiers applied to a set of nine 2D and 3D features extracted from each nodule candidate. These features include (1) the mean gray level of the candidate, (2) the

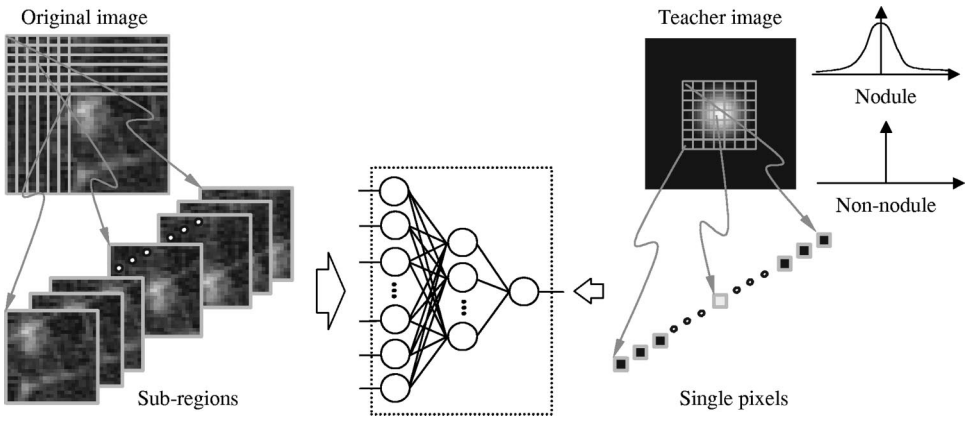


FIG. 2. Architecture and training of the massive training artificial neural network (MTANN). The original image including a nodule or non-nodules is divided pixel by pixel into a large number of overlapping subregions. All pixel values in each of the subregions (e.g., 9×9 matrix size) are entered as input to the MTANN, whereas a pixel value of each single pixel from the teacher image is used as the teacher value. Thus, a large number of subregions and the corresponding single pixels are used for training, e.g., 19×19 subregions and 19×19 single pixels are used for one nodule image or one non-nodule image.

gray-level standard deviation, (3) the gray-level threshold at which the candidate was identified, (4) volume, (5) sphericity, (6) radius of the sphere of equivalent volume, (7) eccentricity, (8) circularity, and (9) compactness.

With our current CAD scheme, the multiple gray-level-thresholding technique initially identified 20 743 nodule candidates in 1057 sections of LDCT images in the training set.[19] Forty-five of 50 nodules were correctly detected. Then a rule-based classifier followed by a series of two linear discriminant classifiers was applied for removal of some false positives, thus yielding a detection of 40 (80.0%) of 50 nodules (from 22 patients) together with 1078 (1.02 per section) false positives.[19] The sizes of the 10 false negative nodules ranged from 5 mm to 25 mm, and the mean diameter was 13.2±6.1 mm. In this study, we used all 50 nodules, the locations of which were identified by the radiologist, and all 1078 false positives generated by our CAD scheme in the training set, for investigating the characteristics of the MTANN and training the MTANN. The use of radiologist-extracted true nodules with computer-generated false positives was intended to anticipate future improvements in the nodule detection sensitivity of our CAD scheme. When a nodule was present in more than one section, the section that included the largest nodule was used. When we applied our current CAD scheme to the test set, a sensitivity of 81.7% (58 of 71 nodules) with 0.98 false positives per section (1726/1765) was achieved. We used the 58 true positives (nodules from 54 patients) and 1726 false positives (non-nodules) for testing the MTANN in a validation test.

III. METHODS

A. Background of ANN and image processing

Recently, in the field of signal processing, nonlinear filters based on a multilayer ANN, called ''neural filters,'' have been studied. In the neural filter, the multilayer ANN is employed as like a convolution kernel. The neural filters can acquire the functions of various linear and nonlinear filters through training. It has been demonstrated that the neural filters can represent an averaging filter, weighted averaging filters, weighted median filters, morphological filters, microstatistic filters, generalized-weighted-order statistical filters, an epsilon filter, and generalized stack filters.[20–23] Suzuki *et al.*[24–30] developed the neural filters for reduction of the quantum mottle in x-ray fluoroscopic and radiographic images, and they reported that the performance of the neural filter was superior to that of the nonlinear filters utilized in medical systems and to that of another, well-known nonlinear filter.

On the other hand, in the field of computer vision, Suzuki *et al.*[31–34] developed a supervised edge detector based on a multilayer ANN, called a ''neural edge detector.'' The neural edge detector can acquire the function of a desired edge detector through training. It has been reported that the performance of the neural edge detector in the detection of edges from noisy images was far superior to that of conventional edge detectors such as the Canny edge detector, the Marr–Hildreth edge detector, and the Huckel edge detector.[31,32] In its application to the contour extraction of the left ventricular cavity in digital angiography, it has been reported that the neural edge detector can accurately replicate the subjective edges traced by cardiologists.[33,34]

B. Architecture of massive training artificial neural network (MTANN)

We are extending the neural filters and the neural edge detector to accommodate various image-processing and pattern-recognition tasks, and we shall call this technique a massive training artificial neural network (MTANN). The architecture and the training method of the MTANN are shown in Fig. 2. The MTANN consists of a modified multilayer ANN, which can directly handle input gray levels and output gray levels. In the MTANN, image processing or pattern recognition is performed by scanning of an image with the modified ANN in which the activation functions of the units in the input, hidden, and output layers are a linear, a sigmoid, and a linear function, respectively. The MTANN employs a linear 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,[35,36,32] for example (see Appendix for theoretical consideration).

The pixel values of the original images are normalized first such that −1000 HU (Hounsfield units) is zero and 1000 HU is one. The pixel values in a local window R_S are input

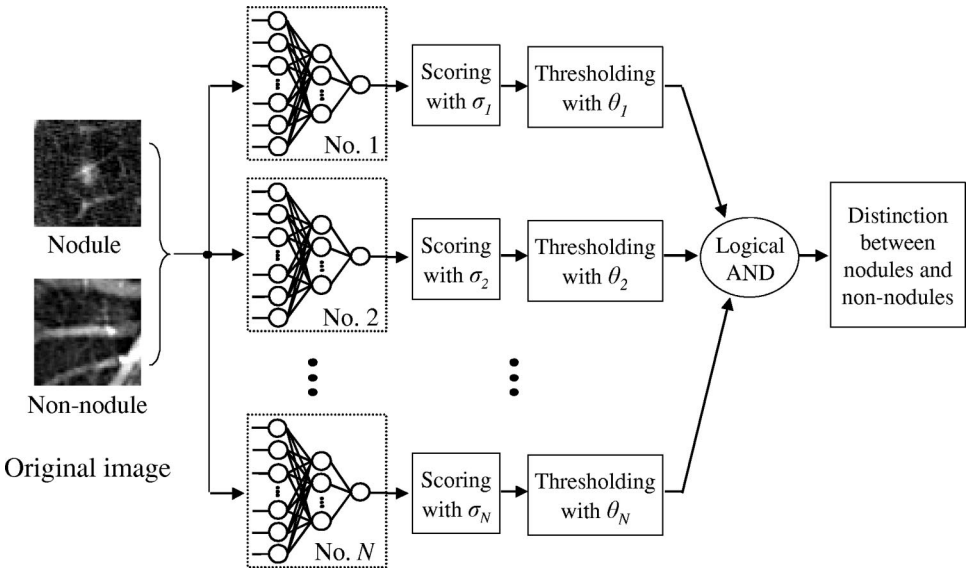


FIG. 3. Architecture of the multiple massive training artificial neural network (Multi-MTANN). Each MTANN is trained by using a different type of non-nodule, but with the same nodules. Each MTANN acts as an expert for distinction between nodules and a specific type of non-nodule. The output of each MTANN is integrated by the logical AND operation.

to the MTANN so that the inputs to the MTANN are a normalized pixel value $g(x,y)$ of an original image, and spatially adjacent normalized pixel values. The output of the MTANN is a continuous value, which corresponds to the center pixel in the local window, represented by

$$f(x,y)=NN\{\mathbf{I}(x,y)\}=NN\{g(x-i,y-j)|i,j\in R_S\}, \quad (1)$$

where $f(x,y)$ denotes the estimate for the teacher value, x and y are the indices of coordinates, $NN\{\cdot\}$ is the output of the modified multilayer ANN, $\mathbf{I}(x,y)$ is the input vector to the modified multilayer ANN, the elements of which are the normalized pixel values in the local window R_S, $g(x,y)$ is the normalized pixel value, and R_S is the local window of the modified multilayer ANN. Note that only one unit is employed in the output layer. The teacher values and thus the outputs of the MTANN need to be changed according to its application. When the task is the distinction between nodules and non-nodules, the output would be interpreted as the ''likelihood of being a nodule.''

All pixels in an image may be entered as input by scanning of the entire image with the MTANN. The local window of the MTANN must be shifted pixel-by-pixel throughout the image. The MTANN can be designed by training such that the input images are converted to the teacher images. The universal approximation property of a multilayer ANN[37,38] guarantees diverse capabilities of the MTANN; i.e., because it has been shown theoretically that a multilayer ANN can realize any continuous mapping approximately, the MTANN can realize through a filtering operation some image-processing and pattern-recognition techniques, including high-pass, low-pass, and band-pass filtering, noise reduction, edge enhancement, edge detection, interpolation, pattern matching, object enhancement, object recognition, aspects of the wavelet transform, aspects of Fourier-based texture analysis, and segmentation. For example, the modified multilayer ANN-based kernel can act as an averaging operation, gradient operation, Laplacian operation, linear and nonlinear interpolation functions, a wavelet function, part of a sinusoidal function, etc. Thus, the MTANN would have a high potential for solving existing problems in CAD as well as image processing and pattern recognition.

In order to distinguish between nodules and various types of non-nodules, we extended the capability of the single MTANN and developed a multiple MTANN (Multi-MTANN). 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 using a different type of non-nodule, but with the same nodules. Each MTANN acts as an expert for distinction between nodules and a specific type of non-nodule, e.g., MTANN No. 1 is trained to distinguish nodules from false positives caused by medium-sized vessels; MTANN No. 2 is trained to distinguish nodules from soft-tissue-opacity false positives caused by the diaphragm; and so on. A scoring method is applied to the output of each MTANN, and then thresholding of the score from each MTANN is performed for distinction between nodules and the specific type of non-nodule. The output of each MTANN is then integrated by the logical AND operation. If each MTANN can eliminate the specific type of non-nodule with which the MTANN is trained, then the Multi-MTANN will be able to reduce a larger number of false positives than does a single MTANN.

C. Training of MTANN

In order to learn the relationship between the input image and the teacher image, the MTANN is trained with a set of input images and the teacher images by adjusting the weights between layers. The error to be minimized by training is defined by

$$E=\frac{1}{2P}\sum_p \{T^{(p)}-f^{(p)}\}^2, \quad (2)$$

where p is a training pixel number, $T^{(p)}$ is the pth training pixel in the teacher images, $f^{(p)}$ is the pth training pixel in the output images, and P is the number of training pixels.

The MTANN is trained by the modified back-propagation (BP) algorithm,[35] which was derived for the structure described above, 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.[39,40] In the modified BP algorithm, the correction of the weight between the mth unit in the hidden layer and the unit in the output layer is represented by

$$\Delta W_m^O = -\eta \cdot \delta \cdot O_m^H = -\eta (T-f) O_m^H, \tag{3}$$

where η is the learning rate, O_m^H is the output of the mth unit in the hidden layer, and δ is the delta of the delta rule.[39,40] By use of the delta, the corrections of any weights can be derived in the same way as in the derivation of the BP algorithm.

For distinguishing between nodules and non-nodules, the teacher image is designed to contain the distribution for the likelihood of being a nodule, i.e., the teacher image for nodules should contain a certain distribution, the peak of which is located at the center of the nodule, and that for non-nodules should contain zeros. As the distance increases from the center of the nodule, the likelihood of being a nodule decreases; therefore, we use a two-dimensional Gaussian function with standard deviation σ_T at the center of the nodule as the distribution for the likelihood of being a nodule, where σ_T may be determined as a measure representing the size of nodules.

Figure 2 illustrates the training for an input image that contains a nodule near the center. First, the input image is divided pixel-by-pixel into a large number of overlapping subregions. The centers of consecutive subregions in Fig. 2 differ by just one pixel. The size of the subregion corresponds to that of the local window R_S of the MTANN. All pixel values in each of the subregions are entered as input to the MTANN, whereas one pixel from the teacher image is entered into the output unit in the MTANN as the teacher value. This single pixel is chosen at the location in the teacher image that corresponds to the center of the input subregion. By presenting each of the input subregions together with each of the teacher single pixels, the MTANN is trained. The training set for each nodule or non-nodule image is represented by the following equations:

$$\begin{aligned} \{\mathbf{I}_s(x,y) | x,y \in R_T\} &= \{\mathbf{I}_{s1}, \mathbf{I}_{s2}, \ldots, \mathbf{I}_{sq}, \ldots, \mathbf{I}_{sN_T}\}, \\ \{T_s(x,y) | x,y \in R_T\} &= \{T_{s1}, T_{s2}, \ldots, T_{sq}, \ldots, T_{sN_T}\}, \end{aligned} \tag{4}$$

where s is the image number, R_T is the training region, q is the pair number of the input subregion and the teacher pixel, N_T is the number of pixels in R_T, $\mathbf{I}_s(x,y)$ is the input vector representing the subregion extracted from the sth input image, and $T_s(x,y)$ is the sth teacher image. Thus, a large number of input subregions overlap each other, and the corresponding teacher pixels in the teacher image are used for training. The MTANN is trained with massive training samples to achieve 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. After training, the MTANN is expected to output the highest value when a 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 is a non-nodule.

In the Multi-MTANN, each MTANN is trained independently by use of the same nodules and a different set of non-nodules. First, the false positives (non-nodules) reported by the CAD scheme for lung nodule detection in CT are classified into a number of groups. The number of groups may be determined by the number of different kinds of false positives. Typical non-nodules in each group are selected as training samples for each MTANN separately, whereas typical nodules are selected as training samples for all MTANNs. The input images and the teacher images are used to train each MTANN in the same way as a single MTANN is trained, based on the modified BP algorithm.[35] The individual MTANNs are expected to act as experts for the specific type of non-nodule after training.

D. Scoring of the MTANN output for testing

When an original image for the sth nodule candidate is entered into the nth trained MTANN for testing, the output image for the sth nodule candidate is obtained by scanning of the original image with the trained MTANN. The distinction between a nodule and a non-nodule is determined by use of a score defined from the output image of the nth trained MTANN, described as follows:

$$S_{ns} = \sum_{x,y \in R_E} f_G(\sigma_n; x,y) \times f_{ns}(x,y), \tag{5}$$

where S_{ns} is the score of the nth trained MTANN for the sth nodule candidate, R_E is the region for evaluation, $f_{ns}(x,y)$ is the output image of the nth trained MTANN for the sth nodule candidate where its center corresponds to the center of R_E, $f_G(\sigma_n; x,y)$ is a two-dimensional Gaussian function with standard deviation σ_n where its center corresponds to the center of R_E, and n is the MTANN number in the Multi-MTANN. This score represents the weighted sum of the estimate for the likelihood of the image containing a nodule near the center, i.e., a higher score would indicate a nodule, and a lower score would indicate a non-nodule.

In the Multi-MTANN, the distribution in the output image of each trained MTANN may be different according to the type of non-nodule trained. The output from each trained MTANN is scored independently by use of a two-dimensional Gaussian function with a different standard deviation σ_n. The standard deviations σ_n were determined by use of a training set. The distinction between nodules and the specific type of non-nodule is determined by thresholding the score with a different threshold θ_n for each trained MTANN, because the appropriate threshold for each trained MTANN may be different according to the type of non-nodule trained. The threshold θ_n may be determined by use of a training set so as not to remove any nodules, but eliminate non-nodules as much as possible. The outputs of the expert MTANNs are combined by use of the logical AND operation such that

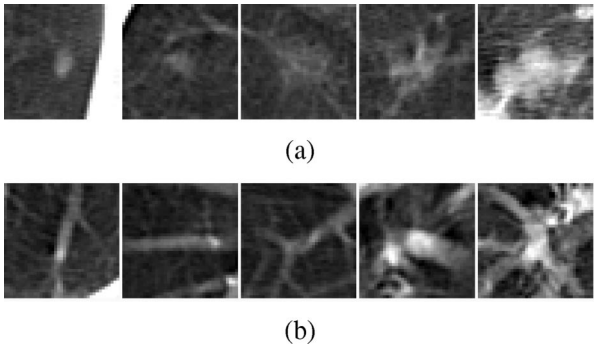


FIG. 4. Illustration of (a) nodules and (b) non-nodules used as training samples for MTANN training.

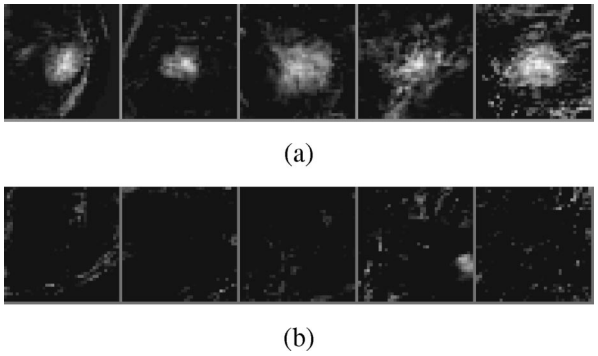


FIG. 5. Output images of the MTANN for nodule candidates of (a) nodules and (b) non-nodules in training samples shown in Figs. 4(a) and 4(b), respectively. These images illustrate the results in which the trained MTANN is applied to the training samples.

each of the trained MTANNs eliminates none of the nodules, but removes some of the specific type of non-nodule for which the MTANN was trained.

IV. RESULTS

The training set in our database consisted of 38 LDCT scans (a total of 1057 LDCT 512×512-pixel images) which included 50 nodules. The 10 nodules and 10 false positives were used as the training samples for the MTANN. Examples of the training samples (a region of 40×40 pixels is displayed as an example) are shown in Fig. 4. One of the coauthors (K.S.) selected 10 typical nodules as training samples from the three categories (pure GGO, mixed GGO, and solid nodule) determined by three radiologists (F.L. and two other experienced chest radiologists) on the basis of the visual appearance of these patterns. The distribution of nodule sizes of training samples is shown in Fig. 1. Six of the 10 nodules were present in one section, two nodules were present in two sections, and two nodules were present in three sections. When a nodule was present in more than one section, the section that included the largest nodule was used. Note that eight of the 10 training nodules were included in 40 nodules detected by our current CAD scheme. Three of the 10 nodules were attached to the pleura, three nodules were attached to vessels, and one nodule was in the hilum. A radiologist (F.L.) classified the false positives reported by our current CAD scheme as four major groups such as small (including peripheral) vessels (40% of false positives), medium-sized vessels (30%), soft-tissue opacities including opacities caused by the partial volume effect between the lung region and the diaphragm (20%), and part of normal structures in the mediastinum, including large vessels in the hilum (10%). Because small (including peripheral) vessels were included in the medium-sized vessel images, we selected medium-sized vessels as the group used for training samples. The radiologist selected 10 vessels with relatively high contrast from the group of the medium-sized vessels, because they are dominant over all medium-sized vessels.

A three-layer structure was employed as the structure of the modified multilayer ANN, because any continuous mapping can be approximately realized by three-layer ANNs.[37,38] The parameters such as the size of the local window of the MTANN R_S, the standard deviation of the two-dimensional Gaussian function σ_T, and the size of the training region in the teacher image R_T were determined empirically based on the training set (see details in the Discussion section). R_S was selected to be 9×9 pixels. The number of units in the hidden layer was set at 25. Thus, the numbers of units in the input, hidden, and output layers were 81, 25, and one, respectively. σ_T was determined as 5.0 pixels, which corresponds approximately to the average diameter of the nodules. R_T was selected to be 19×19 pixels. With the parameters above, the training of the MTANN was performed on 500 000 epochs—one epoch means one training run for one training data set—and converged with a mean absolute error of 11.2%. The training was stopped at 500 000 epochs, because the performance did not increase (see details in the Discussion section). The training took CPU time of 29.8 hours on a PC-based workstation (CPU: Pentium IV, 1.7 GHz), and the time for applying the trained MTANN to nodule candidates was negligibly small.

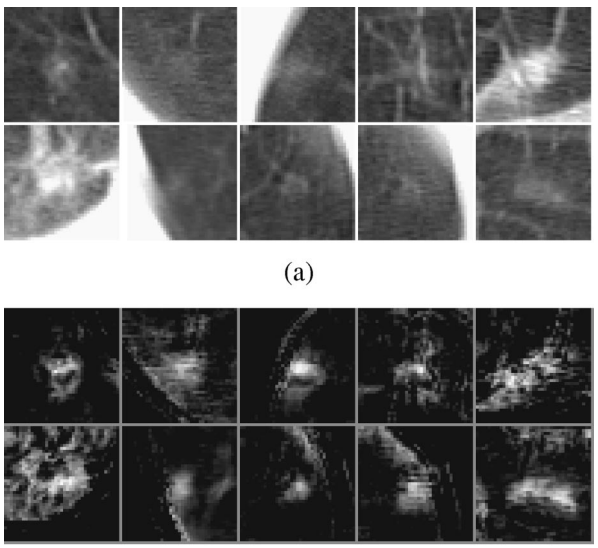
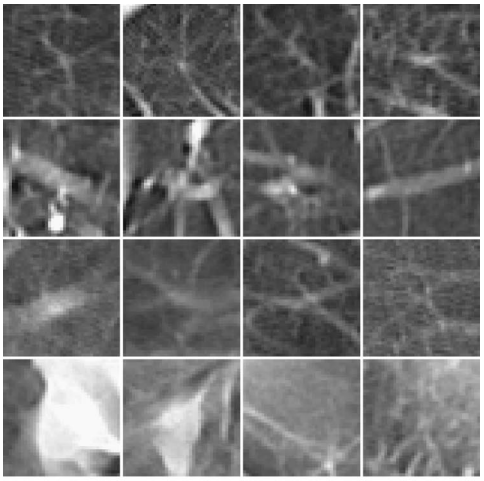
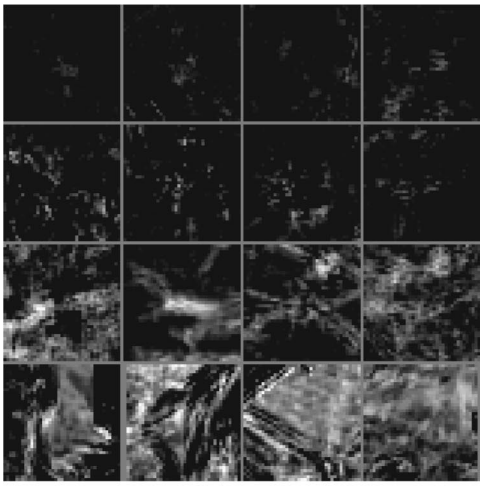


FIG. 6. Illustration of (a) 10 nodules and (b) the corresponding output images of the MTANN for nontraining cases.



(a)



(b)

FIG. 7. Illustration of (a) false positives (non-nodules) and (b) the corresponding output images of the MTANN for nontraining cases. The top, second, third, and fourth rows show typical examples of very small (including peripheral) vessels, medium-sized vessels with relatively high contrast, some other vessels, and some other opacities, respectively. In the third row, the images are medium-sized vessels (relatively large fuzzy vessels), medium-sized vessels (relatively small branching vessels), small (including peripheral) vessels, and peripheral vessels with a light background. In the fourth row, the images are large vessels in the hilum, relatively large vessels with some opacities, soft-tissue opacities caused by the partial volume effect, and some abnormal opacities (focal interstitial opacities), respectively.

The results of applying the trained MTANN to the nodules used for training are shown in Fig. 5. Before we applied the trained MTANN, pixels outside the segmented lung regions reported by our current lung nodule detection scheme[13–16] were set to −1000 HU. The nodules in the output images of the MTANN are represented by light distributions near the center, whereas the output images for false positives (non-nodules) are almost uniformly dark. It is apparent that the distinction between nodules and non-nodules in the output images of the MTANN is superior to that in the original images. Therefore, the MTANN was able to learn important features related to the input images and the teacher images.

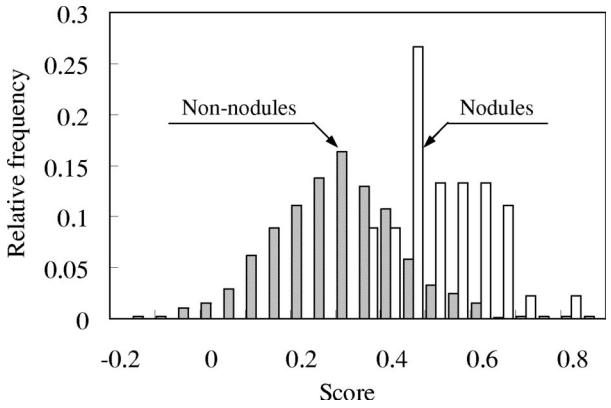


FIG. 8. Histograms of the scores for 40 nodules and 1068 non-nodules, which were different from the ten nodules and ten non-nodules used as training samples.

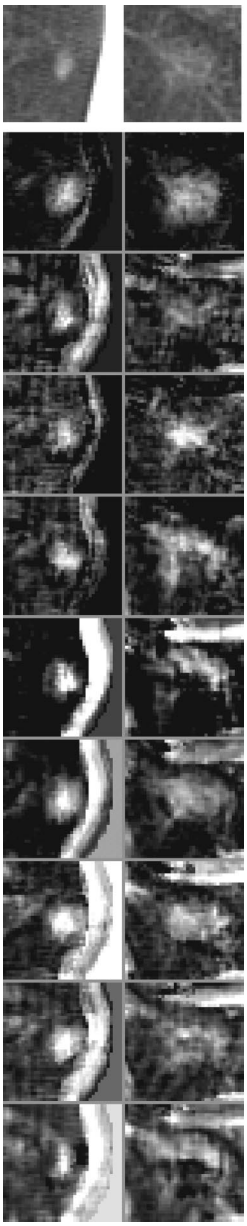


FIG. 9. Illustration of two nodules (top row)used for training the Multi-MTANN and nine corresponding pairs of output images from the nine MTANNs trained separately with different types of non-nodules, as shown in Fig. 10.

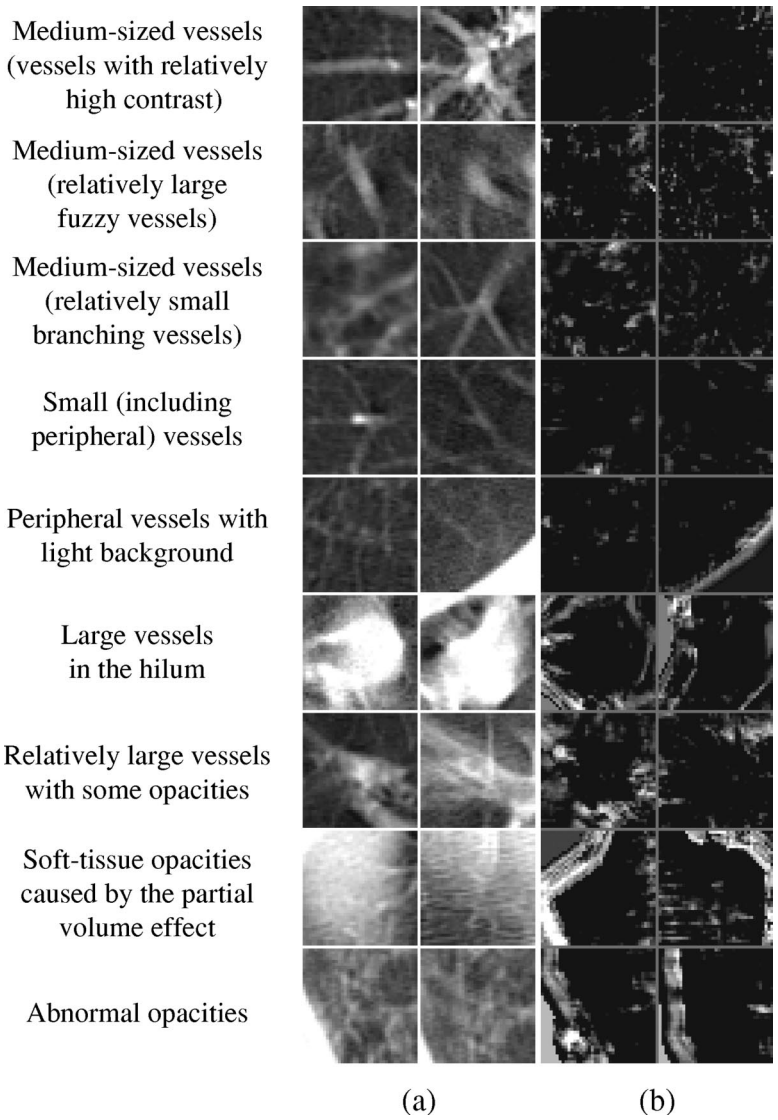


FIG. 10. Illustration of (a) nine sets of non-nodules (two examples in each group) and (b) the corresponding output images from the nine trained MTANNs for training samples.

The trained MTANN was applied to 1068 false-positive nodule candidates not used for training. The execution time was very short, only 1.4 seconds for 1000 nodule candidates. The results for nontraining cases are shown in Figs. 6 and 7. In the output image of the MTANN for nodules (Fig. 6), the nodules are represented by light distributions as expected. The output images for very small (including peripheral) vessels and medium-sized vessels (with relatively high contrast) are almost uniformly dark, as shown in Fig. 7. Because 70% of false positives are small (including peripheral) and medium-sized vessels, we can reduce a large number of false positives by using the output images of the MTANN. However, the output images for other vessels and other opacities such as medium-sized vessels (relatively large fuzzy vessels), medium-sized vessels (relatively small branching vessels), small (including peripheral) vessels, peripheral vessels with a light background, large vessels in the hilum, relatively large vessels with some opacities, soft-tissue opacities caused by the partial volume effect between peripheral vessels and the diaphragm, and some abnormal opacities (focal interstitial opacities) are relatively light. Thus, this trained MTANN was not effective for those false positives, because they are quite different from the vessels used for training in terms of the appearance of patterns such as the contrast, size, gray-level shape, geometric shape, the background gray level, and texture.

The scoring method was applied to the output images of the MTANN. The standard deviation of the Gaussian function for scoring was determined as $\sigma_1 = 4.0$ by use of empirical analysis based on the training set (see details in the Discussion section). We used an R_E of 25×25 pixels. Figure 8 shows the distribution of the scores for the 40 nodules and 1068 non-nodules used for testing; these were different from the 10 nodules and 10 non-nodules used for training. Although the two distributions overlap, it is possible to distinguish a large number of non-nodules from nodules. We determined the threshold so as not to eliminate any nodules, but to remove non-nodules as much as possible. We could remove 66% (706/1068) of false positives (non-nodules).

In order to eliminate the remaining false positives, we prepared training samples for the Multi-MTANN. The radiologist (F.L.) classified the remaining false positives (non-

nodules) reported by the single MTANN into seven groups such as medium-sized vessels, small (including peripheral) vessels, parts of normal structures including large vessels in the hilum, vessels with some opacities, opacities caused by the partial volume effect, abnormal opacities, and other opacities. Two major groups were divided into subgroups based on the visual appearance of patterns. The group of medium-sized vessels was divided into two subgroups such as relatively large fuzzy vessels and relatively small branching vessels. The group of small vessels was divided into two subgroups such as small (including peripheral) vessels and peripheral vessels with a light background. One of the coauthors (K.S.) selected 10 representative non-nodules from each of the groups or the subgroups except the group of other opacities as the training samples for each MTANN; thus, the Multi-MTANN employed nine MTANNs. The same 10 nodules were used as training samples for all nine MTANNs. Therefore, 10 nodules and 90 non-nodules were used for training the Multi-MTANN. Examples of the training samples are shown in Figs. 9 (top row) and 10(a). The single MTANN trained with medium-sized vessels (with relatively high contrast) was used as MTANN No. 1. Non-nodules for the training of MTANN from No. 1 to No. 5 ranged from medium-sized vessels to small (peripheral) vessels. Non-nodules for the training of MTANN from No. 6 to No. 9 were large vessels in the hilum, relatively large vessels with some opacities, soft-tissue opacities caused by the partial volume effect between peripheral vessels and the diaphragm, and some abnormal opacities (focal interstitial opacities), respectively. Each MTANN was trained in the same way as a single MTANN.

The results of applying each of the trained MTANNs to the false-positive nodule candidates and the nodules used as training samples are shown in Figs. 9 and 10(b). The output images of the MTANNs corresponding to nodules demonstrate light distributions near the center, whereas the output images corresponding to false positives (non-nodules) are relatively dark.

The trained Multi-MTANN was applied to the 988 false-positive nodule candidates and 40 nodules not used during training (i.e., the cases were different from the training cases of 10 nodules and 90 non-nodules). The results for nontraining cases are shown in Figs. 11 and 12(b). As shown in Fig. 11, the output images of MTANNs for nodules are represented by light distributions. The output images for false positives are relatively dark around the center, as shown in Fig. 12(b). The output images for large vessels in the hilum (sixth row), soft-tissue opacities (eights and ninth rows), and abnormal opacities (ninth row) are also dark, whereas the individual MTANNs trained for vessels (MTANN from No. 1 to No. 5) were not effective for these false positives.

The scoring method was applied to the output images of the individually trained MTANNs, where σ_n was determined empirically based on the training set to be within the range from 0.5 to 5.5 for the different MTANNs. The performance of the individual MTANNs was evaluated by free-response receiver operating characteristic (FROC) curves,[41] as shown in Fig. 13. The FROC curve expresses a classification sensitivity as a function of the number of false positives per section at a specific operating point, which is determined by the threshold θ_n. With the single MTANN (MTANN No. 1), we can achieve a classification sensitivity of 100% (40/40) with 0.36 false positives per section. The number of false positives per section is defined by

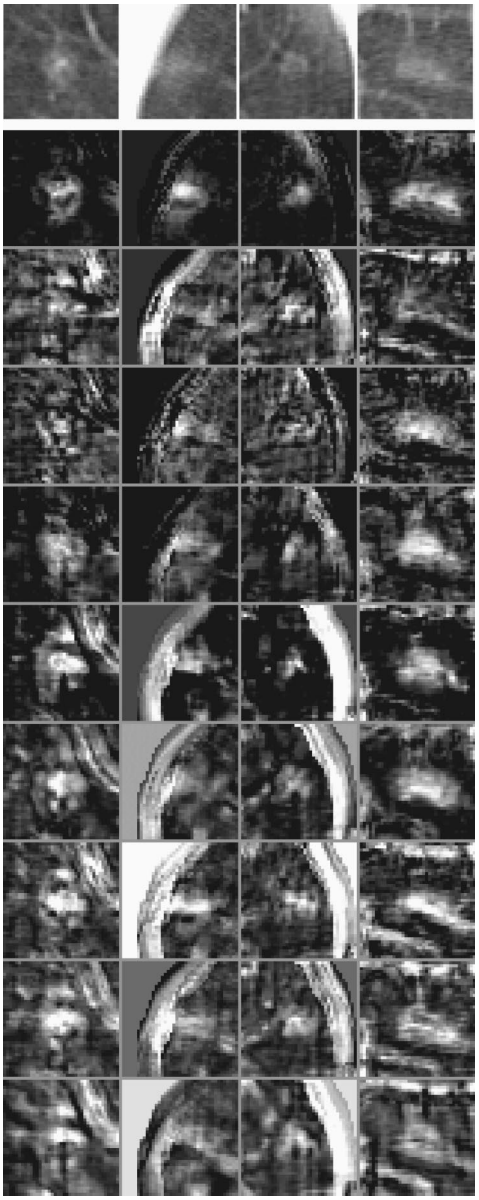


FIG. 11. Illustration of four nodules (nontraining cases) in the top row and nine corresponding sets of output images of the nine trained MTANNs.

$$\mathrm{FPS}=\frac{\mathrm{RFP}}{\mathrm{SEC}\times\left(\dfrac{\mathrm{TFP\text{-}FPT}}{\mathrm{TFP}}\right)}, \tag{6}$$

where RFP is the number of remaining false positives after application of a false-positive reduction method, SEC is the number of section images in the complete database of CT scans, TFP is the total number of false positives reported by the CAD scheme, and FPT is the number of false positives used as training samples, so that the fact that some false

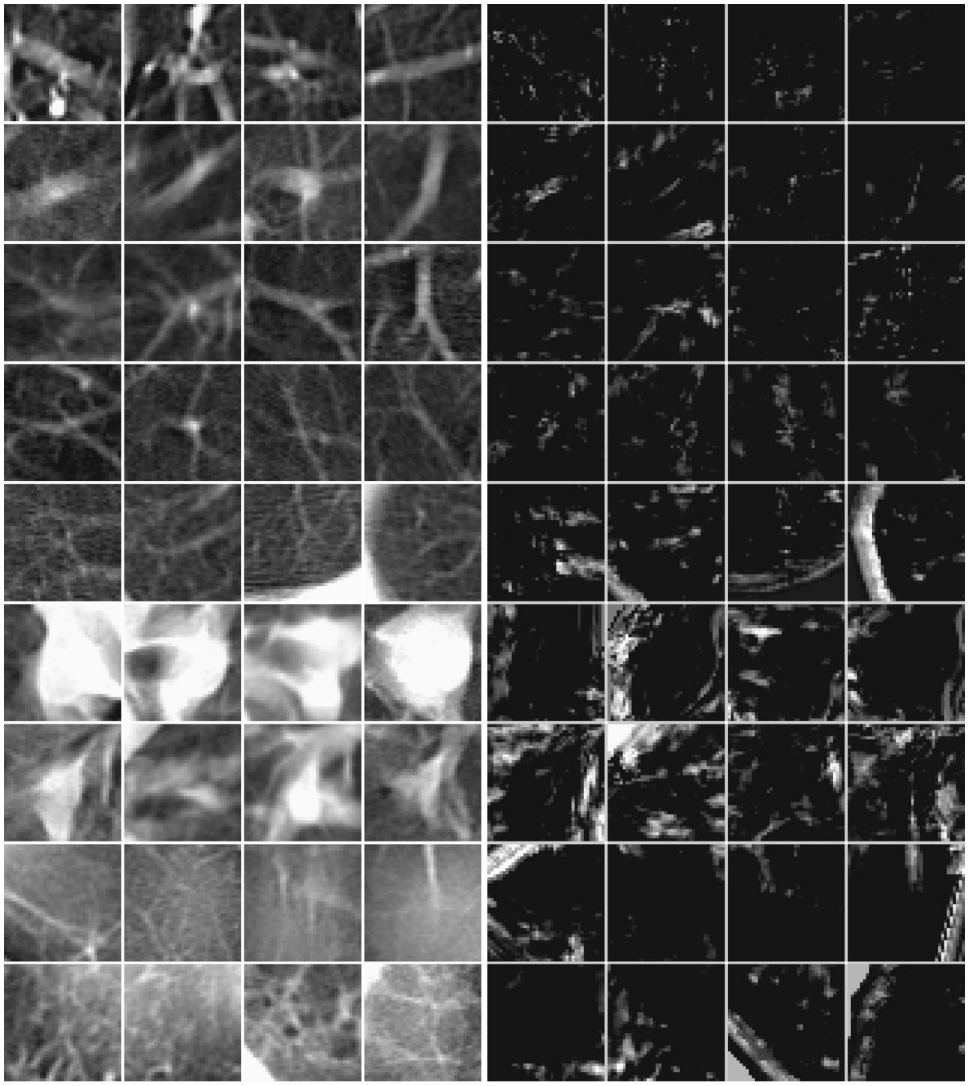
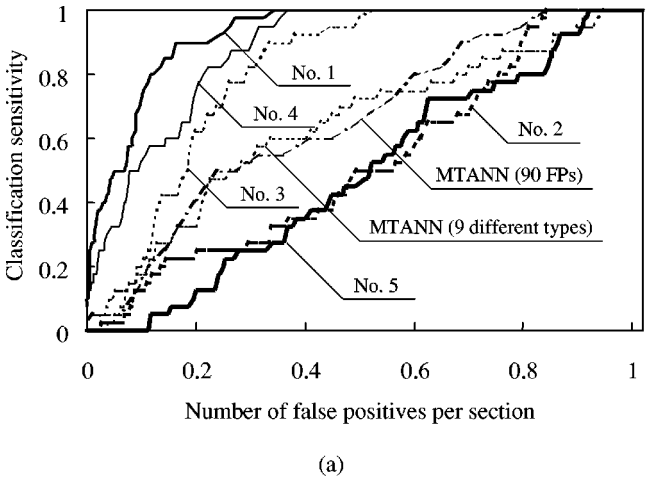


FIG. 12. Illustration of (a) nine sets of non-nodules (four examples in each group) and (b) the corresponding output images from the nine trained MTANNs for nontraining cases. The left-most images from the second to the ninth rows correspond to the images in the third and fourth rows in Fig. 7, for which MTANN No. 1 was not effective.

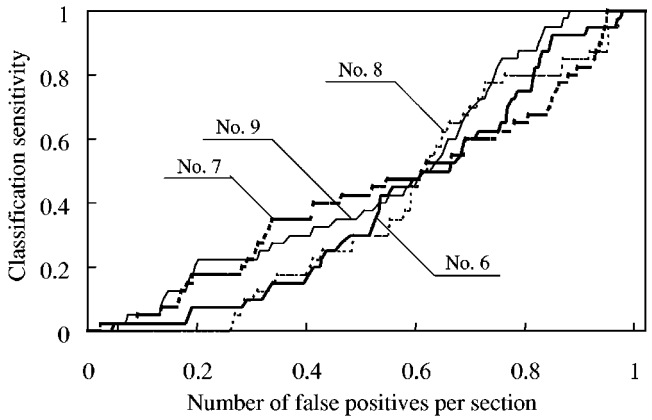
positives were used to train the MTANN does not artificially lower the overall false-positive rate. With the single MTANN (MTANN No. 1), the false-positive rate of our current scheme could be improved from 1.02 to 0.36 false positives per section, while maintaining the current sensitivity. Note that 38 out of 50 nodules used in this study were missed by radiologists.

The performance of the MTANNs varied considerably, because the FROC curves were obtained by testing the MTANNs with all nontraining false positives including various types of non-nodules. The MTANNs trained with dominant false positives such as medium-sized and small (including peripheral) vessels seem to have a better overall performance for all false positives, because the number of medium-sized and small (including) peripheral vessels was the largest of all false positives. The performance of MTANN No. 2 trained with relatively large fuzzy medium-sized vessels was lower among the MTANNs trained with medium-sized vessels (Nos. 1–3), because MTANN No. 2 was not effective for the vessels with high contrast and small vessels, which were dominant over all vessels. The performance of MTANN No. 5 trained with peripheral vessels with a light background was lower than that of MTANN No. 4 trained with small (including peripheral) vessels, because MTANN No. 5 was not effective for small vessels without a light background, which were dominant over all small vessels.

FROC curves of Multi-MTANNs obtained with various numbers of the individual MTANNs are shown in Fig. 14. The FROC curve was the optimized curve where the threshold θ_n of each MTANN in the Multi-MTANN was determined such that the highest performance was obtained, i.e., the threshold θ_n was determined so as not to remove any nodules in the training set, but to remove non-nodules as much as possible. First, the threshold with which none of the nodules was eliminated was calculated for each MTANN. The threshold with the largest number of eliminating non-nodules among all thresholds was used for plotting the FROC curve first. After each of the thresholds was used once for plotting the FROC curve, the second threshold with



(a)



(b)

FIG. 13. FROC curves of the trained MTANNs for (a) No. 1 to No. 5, and (b) No. 6 to No. 9, for 40 nodules and 988 false positives. See details in the Discussion section for MTANN (90 FPs) and MTANN (nine different types).

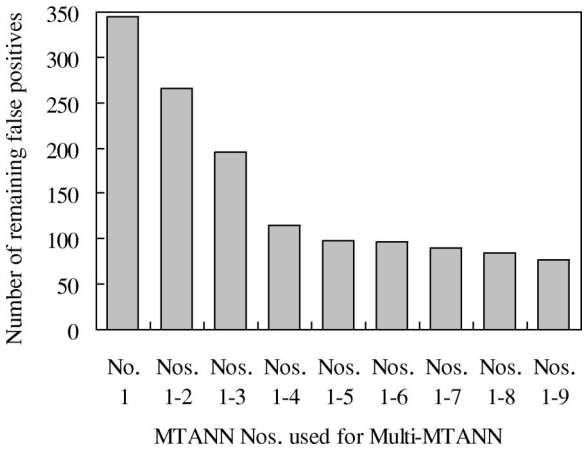


FIG. 15. Number of remaining false positives obtained by Multi-MTANNs at a classification sensitivity of 100%, obtained with various numbers of MTANNs. The number of false positives reported by our current CAD scheme on a database of 38 LDCT scans was reduced from 988 to 76 by use of the Multi-MTANN consisting of nine MTANNs.

which only one nodule was eliminated was calculated. Then, each of the second thresholds was used for plotting the FROC curve in the same manner. These procedures were performed repeatedly, and then the entire FROC curve was obtained by use of the sequences of thresholds $\theta_n(t)$.

We used MTANN No. 1 as the dominant MTANN because MTANN No. 1 had the highest single performance, as shown in Fig. 13. When the Multi-MTANN employed nine MTANNs, 91% (902/988) of false positives (non-nodules) were removed without eliminating any true positives, i.e., we can achieve a classification sensitivity of 100% (40/40) with 0.08 false positives per section. Figure 15 shows the number of false positives that remain after application of the Multi-MTANN at a classification sensitivity of 100% obtained with various numbers of MTANNs.

The test set in our database consisted of 63 LDCT scans (a total of 1765 LDCT images), and contained 71 nodules including 66 biopsy-confirmed primary cancers. The single

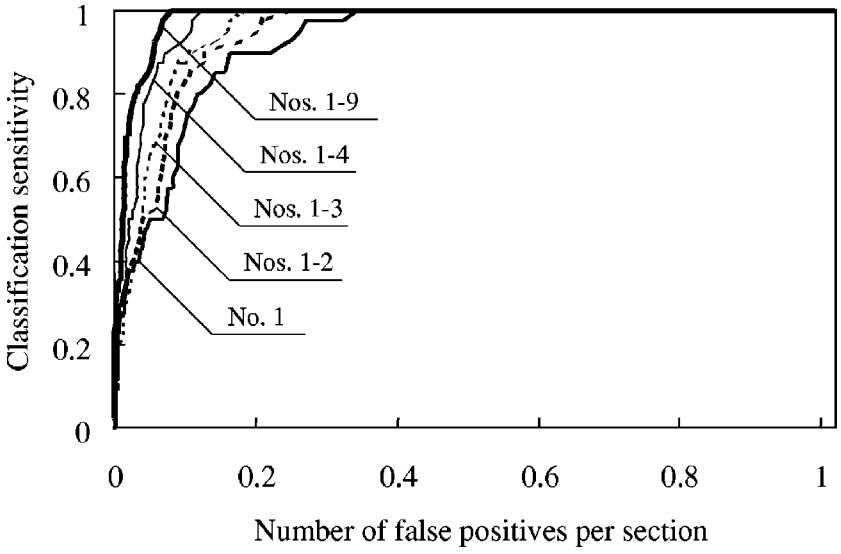


FIG. 14. FROC curves of the Multi-MTANNs obtained with various numbers of the individual MTANNs for 40 nodules and 988 false positives. FROC curve of the Multi-MTANN consisting of nine MTANNs indicates 100% classification sensitivity and a reduction in the false-positive rate from 1.02 to 0.08 per section.

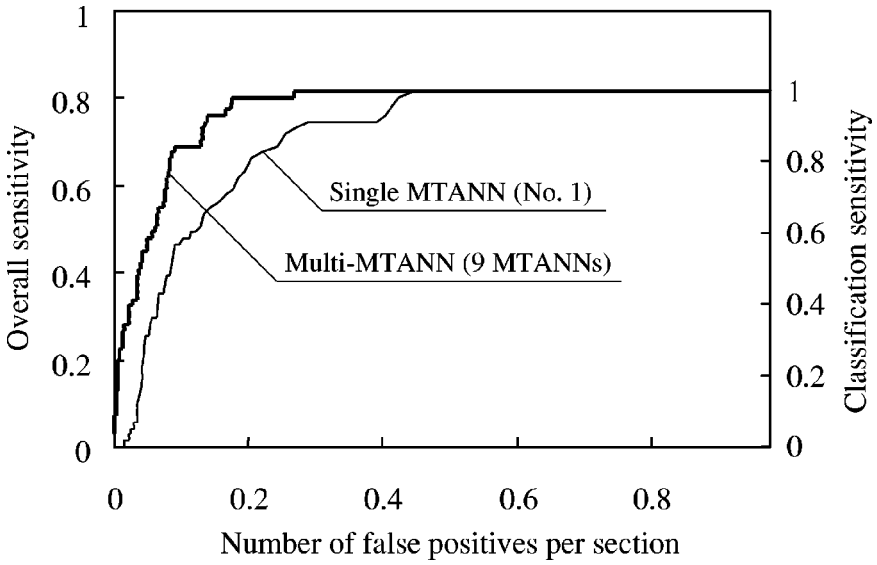


FIG. 16. FROC curves of the single MTANN (MTANN No. 1) and the Multi-MTANN consisting of nine MTANNs for the test set consisting of 57 true positives (nodules) and 1726 false positives (non-nodules) in a validation test. FROC curve of the Multi-MTANN indicates 80.3% overall sensitivity and a reduction in the false-positive rate from 0.98 to 0.18 per section.

MTANN (MTANN No. 1) and the Multi-MTANN which employed nine MTANNs were applied to the 58 true positives (nodules) and 1726 false positives (non-nodules), which were reported by our current CAD scheme for the test set. Note that none of the parameters of the single MTANN and the Multi-MTANN were changed. The FROC curves of the single MTANN and the Multi-MTANN in a validation test are shown in Fig. 16. The FROC curve of the Multi-MTANN was plotted by use of the sequences of thresholds $\theta_n(t)$, which were obtained by generating the optimized FROC curve for the training set. By using the single MTANN, we can remove 54% (938/1726) of false positives (non-nodules) without eliminating any true positives (nodules), i.e., a classification sensitivity of 100% (58/58). The number of eliminated non-nodules increased as the number of MTANNs in the Multi-MTANN increased. When we used the Multi-MTANN including nine MTANNs, 83% (1424/1726) of non-nodules were removed with a reduction of one true positive, i.e., a classification sensitivity of 98.3% (57 of 58 nodules). Note that the performance for the cases in which 23 cases from the same patients as those in the training set were excluded was almost the same, i.e., one true positive was missed at 0.17 false positive per section. Therefore, by using the Multi-MTANN, the false-positive rate of our current CAD scheme was improved from 0.98 to 0.18 false positives per section (from 27.4 to 4.8 per patient) at an overall sensitivity of 80.3% (57/71).

V. DISCUSSION

In order to gain insight into the MTANN, we conducted some experiments with the single MTANN (corresponds to MTANN No. 1). Figure 17(a) shows the effect of the number of training samples on the performance for nontraining cases. The performance was evaluated by using receiver operating characteristic (ROC) analysis.[42,43] One of the authors (K.S.) selected nodules and non-nodules as training samples from the training set. The non-nodules were selected from the group of medium-sized vessels (with relatively high contrast). Note that a set with a larger number of training samples included a set with a smaller number of training samples. There was little increase in the Az value (the area under the ROC curves)[44] when the number of training samples was greater than 20 (10 nodules and 10 non-nodules). This is the reason for using 20 training samples for the MTANN.

Because diagnostic radiology is progressing rapidly as technology advances, a timely development of CAD schemes for diagnostic radiology is important. However, it is difficult for us to obtain a large number of training abnormal cases, particularly for a CAD scheme for diagnosis with a new modality such as a lung cancer screening with CT. The MTANN was able to be trained with such a small number of training samples. The key of this high generalization ability might be due to the division of one nodule image into a large number of subregions. We treated the distinction between nodules and non-nodules as an image-processing task, in other words, as a highly nonlinear filter that performs both nodule enhancement and non-nodule suppression. This allowed us to train the MTANN not on a case basis, but on a subregion basis. The results might suggest that there are some consistent features of nodules in the local window.

In order to gain insight into such a high generalization ability of the MTANN, we investigated the effect of the number of training subregions, i.e., the size of the training region R_T, on the performance for nontraining cases consisting of 40 nodules and 1068 non-nodules. Figure 17(b) shows the Az value when the number of training subregions is varied. The results show that the performance of the MTANN decreased as the number of training subregions decreased. However, there was no increase in Az value when the size of the training region R_T was increased from 19×19 to 25×25. This is the reason for employing 19×19 as the size of the training region R_T. This result suggests that the reason for the high generalization ability of the MTANN is related to the large number of training subregions used. It should be noted that the low performance with a small number of train-

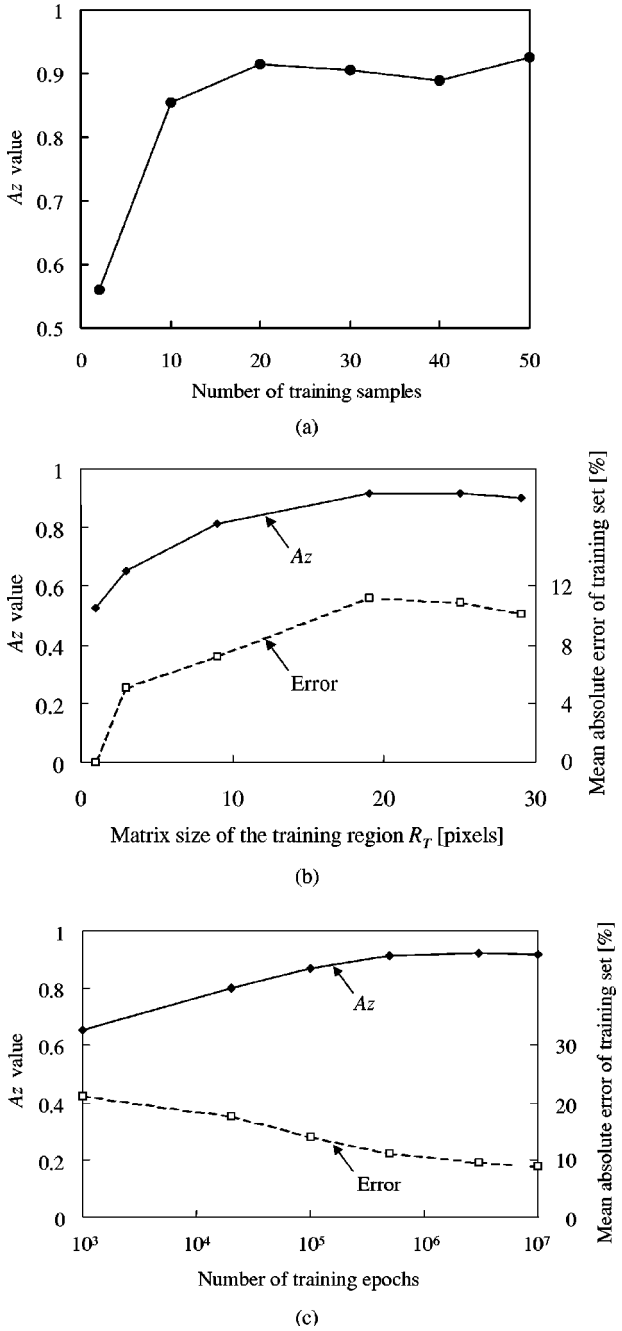


FIG. 17. Effect of (a) the number of training samples, (b) the number of training subregions, and (c) the number of training epochs on the performance for nontraining cases.

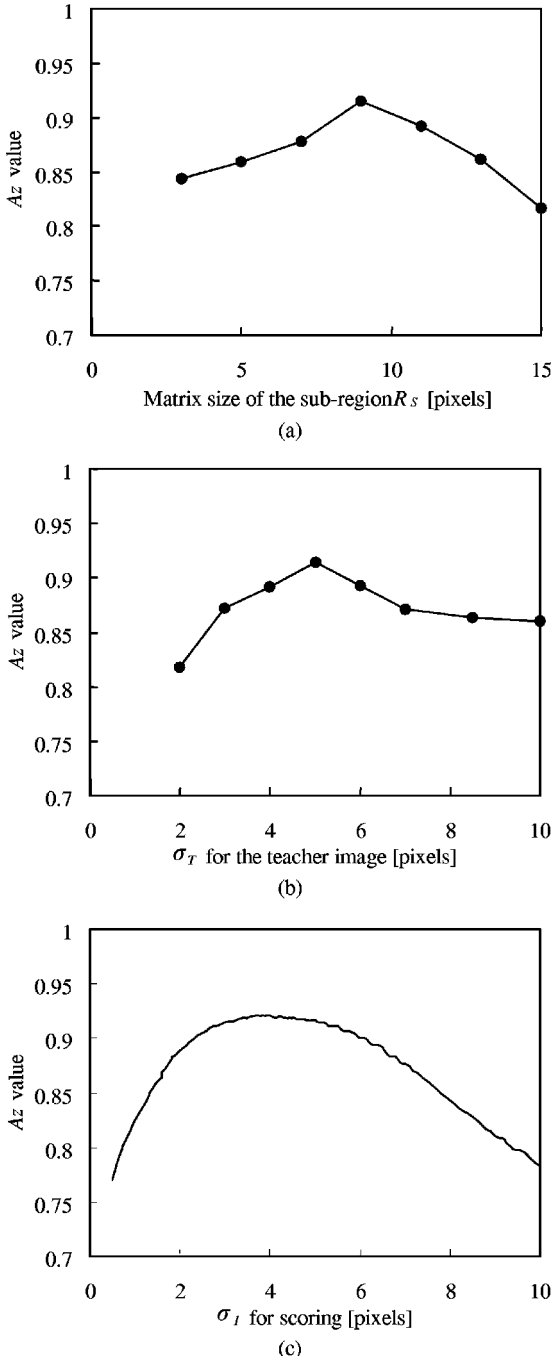


FIG. 18. Effect of (a) the matrix size of the subregion R_S, (b) the standard deviation σ_T of the two-dimensional Gaussian function in the teacher image, and (c) the standard deviation σ_n of the two-dimensional Gaussian function for scoring on the performance for nontraining cases.

ing subregions was not caused by a large training error, as shown in Fig. 17(b).

Figure 17(c) shows the effect of the number of training epochs on the performance. There was little increase in Az value when the number of training epochs was greater than 500 000. This is the reason for determining the condition for stopping the training. Note that overtraining was not seen over 500 000 epochs.

We investigated the effect of the parameter change on the performance. The parameters of the MTANN such as the matrix size of the subregion (local window) R_S, the standard deviation σ_T of the two-dimensional Gaussian function in the teacher image, and the standard deviation σ_1 of the two-dimensional Gaussian function for scoring (for MTANN No. 1) were changed, and the performance for nontraining cases in the training set was obtained, as shown in Fig. 18.

In order to estimate roughly the number of units in the hidden layer required, a method for designing the optimal structure of an ANN[45–47] was applied to the trained MTANN. The method is a sensitivity-based pruning method, i.e., the sensitivity to the training error was calculated when

a certain unit was removed virtually, and the unit with the minimum training error was removed first. The redundant units in the hidden layer were removed on the basis of the effect of removing each unit on the training error, and then the MTANN was retrained to recover the potential loss due to this removal. Each process was performed alternately, resulting in a reduced structure where redundant units were removed. As a result, the optimal number of units in the hidden layer was determined as 22 units.

In order to compare the capability of the single MTANN with that of the Multi-MTANN, we conducted experiments to train the MTANN with various types of non-nodules. We trained the MTANN with the same training samples as used by the Multi-MTANN, i.e., 10 nodules and 90 non-nodules representing nine different non-nodule types. In addition, we trained the MTANN with 10 nodules and nine non-nodules from each of nine non-nodule types (i.e., nine left images in Fig. 10 were used). The results are shown in Fig. 13(a). The performance of the Mulit-MTANN was higher than that of both MTANNs. This result suggests that the capability of the Multi-MTANN was superior to that of the single MTANN.

We examined the performance of the MTANN which is directly applied to the false positives reported by the multiple gray-level-thresholding technique in our current CAD scheme, instead of a combination of the rule-based and linear discriminant classifiers and the MTANN in this study. The majority of the false positives reported by the multiple gray-level-thresholding technique were relatively large vessels whose contrast was relatively high, compared to those reported by our current CAD scheme including rule-based and linear discriminant classifiers. We applied MTANN No. 1 to 20,743 nodule candidates including 45 nodules identified by the multiple gray-level-thresholding technique. We achieved 5.87 false positives per section at a classification sensitivity of 100%, i.e., an overall sensitivity of 90%. At an overall sensitivity of 80%, 1.85 false positives per section were achieved. The false positives eliminated by the rule-based and linear discriminant classifiers were different from those eliminated by the MTANN, although some of them overlapped. A combination of the rule-based and linear discriminant classifiers and the MTANN rather than the MTANN alone might be useful for distinction between nodules and non-nodules in a CAD scheme.

VI. CONCLUSION

We developed a pattern-recognition technique based on an artificial neural network, called a massive training artificial neural network (MTANN), and developed an extended version of the MTANN, called a multiple MTANN (Multi-MTANN), for reduction of false positives in the computerized detection of lung nodules in low-dose CT images. The MTANN and the Multi-MTANN will be useful for substantially reducing the number of false positives in CAD schemes for lung nodule detection.

ACKNOWLEDGMENTS

The authors are grateful to Hiroyuki Abe, M.D., for his clinical advice, Qiang Li, Ph.D., Junji Shiraishi, Ph.D., Yoshikazu Uchiyama, Ph.D., Heber MacMahon, M.D., Maryellen Giger, Ph.D., Roger Engelmann, M.S., and Chisako Muramatsu for their valuable suggestions, Charles E. Metz, Ph.D., for the use of the LABROC5 program, and Elisabeth F. Lanzl for improving the paper. This work was supported by USPHS Grants Nos. CA62625 and CA83908. K. Doi and S. G. Armato are shareholders of R2 Technology, Inc., Sunnyvale, CA. K. Doi is a shareholder of Deus Technologies, Inc., Rockville, MD. It is the policy of the University of Chicago that investigators disclose publicly actual or potential significant financial interest that may appear to be affected by research activities.

APPENDIX

The MTANN consists of the modified multilayer ANN which employs a linear function instead of a sigmoid function as the activation function of the unit in the output layer. In order to clarify the basic property of the modified multilayer ANN and the modified BP algorithm, we considered the relationship between the modified multilayer ANN and the original multilayer ANN theoretically. As for the structure, we can understand easily that it is difficult for the original multilayer ANN to output values near one and zero, whereas the modified multilayer ANN can output all values equally. In the modified BP algorithm, the correction of the weight between the unit in the hidden layer and the unit in the output layer is represented by

$$\Delta W_m^O = -\eta \frac{\partial E}{\partial O^O} \frac{\partial O^O}{\partial X} \frac{\partial X}{\partial W_m^O} = -\eta \frac{\partial E}{\partial O^O} f_L'(X) O_m^H = -\eta \frac{\partial E}{\partial O^O} O_m^H, \tag{A1}$$

where O^O is the output of the unit in the output layer, X is the input value to the activation function, and f_L' is the derivative of the linear function where the slope coefficient is assumed to be 1.0. On the other hand, the correction of the weight in the original BP algorithm is represented by

$$\Delta W_m^{OC} = -\eta \frac{\partial E}{\partial O^O} f_S'(X) O_m^H = -\eta \frac{\partial E}{\partial O^O} O^O(1-O^O) O_m^H, \tag{A2}$$

where f_S' is the derivative of a sigmoid function. Comparing the two equations, we find that the difference is just the derivative of the activation function. Therefore, we can rewrite the right-hand side of Eq. (A2) as the following equation, using η_S:

$$-\eta \frac{\partial E}{\partial O^O} O^O(1-O^O) O_m^H = -\eta_S \frac{\partial E}{\partial O^O} O_m^H. \tag{A3}$$

When the training proceeds, the output of the original multilayer ANN O^O should approach the teacher value T. Therefore, the learning rate of the original BP algorithm can be approximated by

$$\eta_S = \eta \cdot O^O(1-O^O) \approx \eta \cdot T(1-T). \quad \text{(A4)}$$

This equation shows that the learning rate of the original BP algorithm is modulated by the derivative of a sigmoid function, which is 0.5 when the teacher value is 0.5, and is zero when the teacher value is zero or one. In other words, the learning rate of the modified BP algorithm corresponds to that of the original BP algorithm before the modulation. Therefore, in the original BP algorithm, the teacher values of zero and one are never trained, and the training for the teacher value near zero and one converges more slowly. This would affect the convergence characteristic and the output characteristic. Therefore, the modified multilayer ANN with the modified BP algorithm would be suitable for image processing where the teacher values may be continuous values ranging from zero to one, whereas the multilayer ANN is suitable for a classification task where the teacher signal is assigned to a class.

[a]Author to whom correspondence should be addressed; electronic mail: suzuki@uchicago.edu
[b]Also at: Azumi General Hospital, Ikeda, Kitaazumi-gun, Nagano 399-8695, Japan.

[1] A. Jemal, A. Thomas, T. Murray, and M. Thun, "Cancer statistics, 2002," Ca-Cancer J. Clin. **52**, 23–47 (2002).
[2] R. T. Heelan, B. J. Flehinger, M. R. Melamed, M. B. Zaman, W. B. Perchick, J. F. Caravelli, and N. Martini, "Non-small-cell lung cancer: Results of the New York screening program," Radiology **151**, 289–293 (1984).
[3] S. Sone *et al.*, "Mass screening for lung cancer with mobile spiral computed topography scanner," Lancet **351**, 1242–1245 (1998).
[4] M. Kaneko, K. Eguchi, H. Ohmatsu, R. Kakinuma, T. Naruke, K. Suemasu, and N. Moriyama, "Peripheral lung cancer: Screening and detection with low-dose spiral CT versus radiography," Radiology **201**, 798–802 (1996).
[5] C. I. Henschke *et al.*, "Early lung cancer action project: Overall design and findings from baseline screening," Lancet **354**, 99–105 (1999).
[6] J. W. Gurney, "Missed lung cancer at CT: Imaging findings in nine patients," Radiology **199**, 117–122 (1996).
[7] F. Li, S. Sone, H. Abe, H. MacMahon, S. G. Armato III, and K. Doi, "Lung cancers missed at low-dose helical CT screening in a general population: Comparison of clinical, histopathologic, and image findings," Radiology **225**, 673–683 (2002).
[8] S. Yamamoto, I. Tanaka, M. Senda, Y. Tateno, T. Iinuma, T. Matsumoto, and M. Matsumoto, "Image processing for computer-aided diagnosis of lung cancer by CT (LDCT)," Syst. Comput. Japan **25**, 67–80 (1994).
[9] T. Okumura, T. Miwa, J. Kako, S. Yamamoto, M. Matsumoto, Y. Tateno, T. Iinuma, and T. Matsumoto, "Image processing for computer-aided diagnosis of lung cancer screening system by CT (LDCT)," Proc. SPIE **3338**, 1314–1322 (1998).
[10] W. J. Ryan, J. E. Reed, S. J. Swensen, and J. P. F. Sheedy, "Automatic detection of pulmonary nodules in CT," Proc. Computer Assisted Radiology, 1996, pp. 385–389.
[11] K. Kanazawa, M. Kubo, N. Niki, H. Satoh, H. Ohmatsu, K. Eguchi, and N. Moriyama, "Computer assisted lung cancer diagnosis based on helical images," Image Analysis Applications and Computer Graphics: Proc. Int. Computer Science Conf., 1995, pp. 323–330.
[12] M. L. Giger, K. T. Bae, and H. MacMahon, "Computerized detection of pulmonary nodules in computed tomography images," Invest. Radiol. **29**, 459–465 (1994).
[13] S. G. Armato III, M. L. Giger, J. T. Blackbur, K. Doi, and H. MacMahon, "Three-dimensional approach to lung nodule detection in helical CT," Proc. SPIE **3661**, 553–559 (1999).
[14] S. G. Armato III, M. L. Giger, C. J. Moran, J. T. Blackbur, K. Doi, and H. MacMahon, "Computerized detection of pulmonary nodules on CT scans," Radiographics **19**, 1303–1311 (1999).
[15] S. G. Armato III, M. L. Giger, and H. MacMahon, "Analysis of a three-dimensional lung nodule detection method for thoracic CT scans," Proc. SPIE **3979**, 103–109 (2000).
[16] S. G. Armato III, M. L. Giger, and H. MacMahon, "Automated detection of lung nodules in CT scans: Preliminary results," Med. Phys. **28**, 1552–1561 (2001).
[17] J. P. Ko and M. Betke, "Automated nodule detection and assessment of change over time-preliminary experience," Radiology **218**, 267–273 (2001).
[18] S. Sone, F. Li, Z.-G. Yang, S. Takashima, Y. Maruyama, M. Hasagawa, J.-C. Wang, S. Kawakami, and T. Honda, "Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner," Br. J. Cancer **84**, 25–32 (2001).
[19] S. G. Armato III, F. Li, M. L. Giger, H. MacMahon, S. Sone, and K. Doi, "Lung cancer: Performance of automated lung nodule detection applied to cancers missed in a CT screening program," Radiology **225**, 685–692 (2002).
[20] K. Arakawa and H. Harashima, "A nonlinear digital filter using multilayered neural networks," Proc. IEEE Int. Conf. Commun. **2**, 424–428 (1990).
[21] L. Yin, J. Astola, and Y. Neuvo, "A new class of nonlinear filters—neural filters," IEEE Trans. Signal Process. **41**, 1201–1222 (1993).
[22] L. Yin, J. Astola, and Y. Neuvo, "Adaptive multistage weighted order statistic filters based on the back propagation algorithm," IEEE Trans. Signal Process. **42**, 419–422 (1994).
[23] H. Hanek and N. Ansari, "Speeding up the generalized adaptive neural filters," IEEE Trans. Image Process. **5**, 705–712 (1996).
[24] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, "A recurrent neural filter for reducing noise in medical x-ray image sequences," Proc. Int. Conf. Neural Information Processing **1**, 157–160 (1998).
[25] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, "Noise reduction of medical x-ray image sequences using a neural filter with spatiotemporal inputs," Proc. Int. Symp. Noise Reduction for Imaging and Communication Systems, 1998, pp. 85–90.
[26] K. Suzuki, I. Horiba, and N. Sugie, "Training under achievement quotient criterion," *Neural Networks for Signal Processing X* (IEEE, Piscataway, NJ, 2000), pp. 537–546.
[27] K. Suzuki, I. Horiba, and N. Sugie, "Signal preserving training for neural networks for signal processing," Proc. IEEE Int. Symp. Intelligent Signal Processing and Communication Systems **1**, 292–297 (2000).
[28] K. Suzuki, I. Horiba, and N. Sugie, "Neural filter with selection of input features and its application to image quality improvement of medical image sequences," Proc. IEEE Int. Symp. Intelligent Signal Processing and Communication Systems **2**, 783–788 (2000).
[29] K. Suzuki, I. Horiba, and N. Sugie, "Efficient approximation of a neural filter for quantum noise removal in x-ray images," IEEE Trans. Signal Process. **50**, 1787–1799 (2002).
[30] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, "Neural filter with selection of input features and its application to image quality improvement of medical image sequences," IEICE Trans. Inf. Syst. **E85-D**, 1710–1718 (2002).
[31] K. Suzuki, I. Horiba, and N. Sugie, "Edge detection from noisy images using a neural edge detector," *Neural Networks for Signal Processing X* (IEEE, Piscataway, NJ, 2000), pp. 487–496.
[32] K. Suzuki, I. Horiba, and N. Sugie, "Neural edge detector-a good mimic of conventional one yet robuster against noise-," Lect. Notes Comput. Sci. **2085**, 303–310 (2001).
[33] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, "Extraction of the contours of left ventricular cavity, according with those traced by medical doctors, from left ventriculograms using a neural edge detector," Proc. SPIE **4322**, 1284–1295 (2001).
[34] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, "Contour extraction of the left ventricular cavity from digital subtraction angiograms using a neural edge detector," Syst. Comput. Japan **34**, 55–69 (2003).
[35] K. Suzuki, I. Horiba, K. Ikegaya, and M. Nanki, "Recognition of coronary arterial stenosis using neural network on DSA system," Syst. Comput. Japan **26**, 66–74 (1995).
[36] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, "Computer-aided diagnosis system for coronary artery stenosis using a neural network," Proc. SPIE **4322**, 1771–1782 (2001).

[37] K. Funahashi, ''On the approximate realization of continuous mappings by neural networks,'' Neural Networks **2**, 183–192 (1989).
[38] A. R. Barron, ''Universal approximation bounds for superpositions of a sigmoidal function,'' IEEE Trans. Inf. Theory **39**, 930–945 (1993).
[39] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, ''Learning representations of back-propagation errors,'' Nature (London) **323**, 533–536 (1986).
[40] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, ''Learning internal representations by error propagation,'' *Parallel Distributed Processing* (MIT, Cambridge, MA, 1986), Vol. 1, Chap. 8, pp. 318–362.
[41] D. P. Chakraborty and L. H. L. Winter, ''Free-response methodology: Alternate analysis and a new observer-performance experiment,'' Radiology **174**, 873–881 (1990).
[42] C. E. Metz, ''ROC methodology in radiologic imaging,'' Invest. Radiol. **21**, 720–733 (1986).
[43] C. E. Metz, B. A. Herman, and J.-H. Shen, ''Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data,'' Stat. Med. **17**, 1033–1053 (1998).
[44] J. A. Hanley and B. J. McNeil, ''A method of comparing the areas under receiver operating characteristic curves derived from the same cases,'' Radiology **148**, 839–843 (1983).
[45] K. Suzuki, I. Horiba, and N. Sugie, ''Designing the optimal structure of a neural filter,'' *Neural Networks for Signal Processing VIII* (IEEE, Piscataway, NJ, 1998), pp. 323–332.
[46] K. Suzuki, I. Horiba, and N. Sugie, ''A simple neural network pruning algorithm with application to filter synthesis,'' Neural Processing Lett. **13**, 43–53 (2001).
[47] K. Suzuki, I. Horiba, and N. Sugie, ''Simple unit-pruning with gain-changing training,'' *Neural Networks for Signal Processing XI* (IEEE, Piscataway, NJ, 2001), pp. 153–162.