
Computerized Scheme for Automated Detection of Lung Nodules in Low-Dose Computed Tomography Images for Lung Cancer Screening¹

Hidetaka Arimura, PhD,² Shigehiko Katsuragawa, PhD, Kenji Suzuki, PhD, Feng Li, MD, Junji Shiraishi, PhD, Shusuke Sone, MD, Kunio Doi, PhD

Rationale and Objectives. A computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening was developed.

Materials and Methods. Our scheme is based on a difference-image technique for enhancing the lung nodules and suppressing the majority of background normal structures. The difference image for each computed tomography image was obtained by subtracting the nodule-suppressed image processed with a ring average filter from the nodule-enhanced image with a matched filter. The initial nodule candidates were identified by applying a multiple-gray level thresholding technique to the difference image, where most nodules were well enhanced. A number of false-positives were removed first in entire lung regions and second in divided lung regions by use of the two rule-based schemes on the localized image features related to morphology and gray levels. Some of the remaining false-positives were eliminated by use of a multiple massive training artificial neural network trained for reduction of various types of false-positives. This computerized scheme was applied to a confirmed cancer database of 106 low-dose computed tomography scans with 109 cancer lesions for 73 patients obtained from a lung cancer screening program in Nagano, Japan.

Results. This computed-aided diagnosis scheme provided a sensitivity of 83% (91/109) for all cancers with 5.8 false-positives per scan, which included 84% (32/38) for missed cancers with 5.9 false-positives per scan.

Conclusion. This computerized scheme may be useful for assisting radiologists in detecting lung cancers on low-dose computed tomography images for lung cancer screening.

Key Words. Computer-aided diagnosis (CAD); low-dose computed tomography (LDCT); lung cancer screening; difference-image technique.

© AUR, 2004

Acad Radiol 2004; 11:617–629

¹ From the Department of Radiology, The University of Chicago, 5841 S Maryland Ave, Chicago, IL 60637 (H.A., S.K., K.S., F.L., J.S., K.D.); and Azumi General Hospital, Nagano Japan (S.S.). Received January 5, 2003; revision requested February 23; received in revised form February 26, 2003; accepted February 26, 2003. Supported by US Public Health Service grant nos. CA 62625 and CA 98119. K. D. and S. K. are shareholders in R2 Technology, Inc, Los Altos, CA. K.D. is a shareholder of Deus Technologies, Inc, Rockville, MD. **Address correspondence to** H.A. e-mail: arimura@shs.kyushu-u.ac.jp

² Current address: Department of Health Sciences, Kyushu University, Fukuoka 812-8582, Japan.

© AUR, 2004

doi:10.1016/j.acra.2004.02.009

Low-dose helical computed tomography (LDCT) screening is regarded as one of the most promising techniques for early detection of lung cancer (1–5). It has been reported that CT images are superior to chest radiographs for detecting peripheral lung cancers (1). However, it is a difficult and time-consuming task for radiologists to detect subtle lung nodules in a large number of CT slices for lung cancer screening. Thus, a computer-aided diagnostic (CAD) scheme would be useful in assisting radiologists for cancer screening by using LDCT.

A number of investigators (6–14) have attempted to develop CAD schemes for computerized detection of lung nodules by using various methods and techniques. In an early attempt, Yamamoto et al (6) developed a CAD sys-

tem in lung cancer screening with LDCT based on a morphologic filter (Qoit filter), and with this technique Jiang et al (7) reported 100% sensitivity with 0.14 false-positives per slice for 82 cases (2,160 slices), including 21 cancers. Ukai et al (8) developed a prototype CAD system for automated detection of suspicious regions from LDCT images by using a fuzzy clustering method and reported a sensitivity of 95%, although the number of false-positives was not reported. Armato et al (9,10) reported a computerized method for detection of lung nodules in helical CT scans based on 2-dimensional and 3-dimensional analyses of the image data. Armato et al (11) evaluated the performance of the CAD system for a database of LDCT scans with 38 cancers missed in a screening program; the CAD system correctly detected 84% of all cancers with 1.0 false-positive per slice. The database used by Armato et al (11) was a subgroup of our database used for this study. Wormanns et al (12) reported the evaluation of a CAD workstation with an automatic detection algorithm of pulmonary nodules with LDCT in a clinical setting for early detection of lung cancer; the sensitivity for 88 CT examinations was 38% with about 5.8 false-positives per scan. Recently, Gurcan et al (13) reported a preliminary evaluation of a CAD system for lung nodule detection on helical CT images; with their scheme based on weighted k-means clustering segmentation, the sensitivity was 84% with 1.74 false-positives per slice. Brown et al (14) developed an automated system for detecting lung micronodules on thin-section CT images based on 3-dimensional segmentation by use of a model of lung nodules and intrathoracic anatomy; they reported a sensitivity of 100% for nodules (>3 mm in diameter) and 70% for micronodules (≤ 3 mm) with 15 false-positives per scan. However, further efforts would still be required for development of CAD systems for lung cancer screening by use of LDCT with a higher sensitivity and a lower number of false-positives per scan.

Our purpose in this study was to develop a computerized scheme for automated detection of pulmonary nodules in LDCT scans for lung cancer screening. Our CAD scheme was based on a difference-image technique (15–17) developed for enhancing nodule-like objects in chest radiographs, and then selecting initial nodule candidates. Because the purpose of lung cancer screening is to find cancers as early as possible, we used a confirmed cancer database including “missed” cancer scans obtained from a lung cancer screening program for development and testing of the CAD scheme.

MATERIALS AND METHODS

Database

A database of 106 LDCT scans used for this study was acquired on a mobile LDCT scanner (CT-W950SR; Hitachi Medical, Tokyo, Japan) from 73 lung cancer patients obtained in a lung cancer screening program in Nagano, Japan (1996–1999) (2,18). All scans were acquired according to a low-dose protocol at a tube voltage of 120 kV, a tube current of 25 mA or 50 mA, 10 mm collimation, and a 10 mm reconstruction interval at a helical pitch of two. The number of CT slices per scan used for this study was 31 or 33, and this database included 3,292 slices. Each image was 512×512 pixels with a pixel size of 0.586 mm or 0.684 mm, and the number of gray levels was 4,096. Each of the 73 patients whose screening CT data were used in this study had at least one cancer. All cancers were intrapulmonary and were subsequently confirmed to be lung cancers on the basis of results of biopsy or surgery. We excluded cancers larger than 30 mm, and also central cancers, which are endobronchial tumors in or proximal to a segmental bronchus, from an original Nagano database (18).

Because the lesions of 26 cancers had been “missed” one or more times in the screening program, the database included one or more scans for the same patients. This database of 106 scans with 109 confirmed cancers consisted of 68 scans for 68 patients with 71 cancers, which were prospectively detected by radiologists, and also 38 scans for 31 patients with 38 cancers, which were retrospectively identified and considered “missed” because of detection error or interpretation error during clinical interpretation (18). Twenty-three missed cancers identified as detection errors were not mentioned in the radiologists’ reports, and 15 cancers identified as interpretation errors were reported, but misinterpreted. This database also included 22 other nodules, ie, 13 benign nodules confirmed by follow-up diagnostic CT examinations, eight suspicious benign nodules, and one suspicious malignant nodule, where suspicious nodules were based on agreement between two radiologists, but not confirmed by surgery or follow-up CT examinations. Figure 1 shows the distributions of nodule sizes (effective diameters) for the 71 detected cancers ranging from 6 mm to 24 mm with a mean of 14 mm, and 38 missed cancers ranging from 6 mm to 26 mm with a mean of 12 mm. The 131 nodules were found in 226 slices of a total of 3,292 slices, with some of the nodules existing in several slices. If we would in-

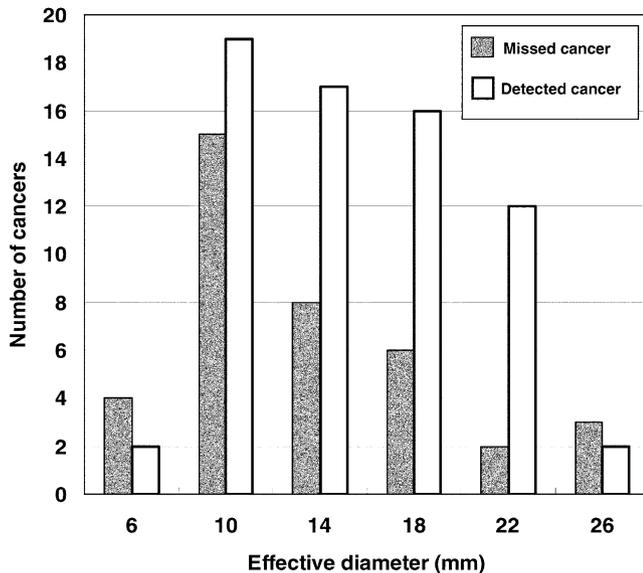


Figure 1. Distributions of nodule sizes for missed cancers and detected cancers in the lung cancer database used in this study.

clude all slices for evaluation, it would become difficult to summarize the overall performance because there would be a mixture of results yielding some “true” positives and some “false” negatives in several slices for one nodule. Therefore, for the purpose of designing rules and evaluating computer detections of nodules in this study, we manually selected one representative CT slice with the largest nodule diameter for each nodule, but we excluded 99 nodules on 99 other slices; this is because the performance of our scheme would be evaluated adequately as long as one nodule in one of the slices would be detected. Consequently, the number of slices, where each nodule existed in one slice, was 127, because four slices included two nodules.

Overall Scheme

Our overall scheme for automated detection of pulmonary nodules on LDCT images is shown in Fig 2. Our CAD scheme is based on a difference-image technique (15–17), by which structures similar to nodules were enhanced, and most of the background normal structures such as small vessels or background noise were suppressed. In this scheme, at first, the left or right lung region was segmented by use of linear discriminant analysis (LDA) (19) on the histogram of CT values for the entire body region, which usually have two main peaks in histograms for lung regions and other tissue regions. By use of LDA, a threshold CT level was automatically determined

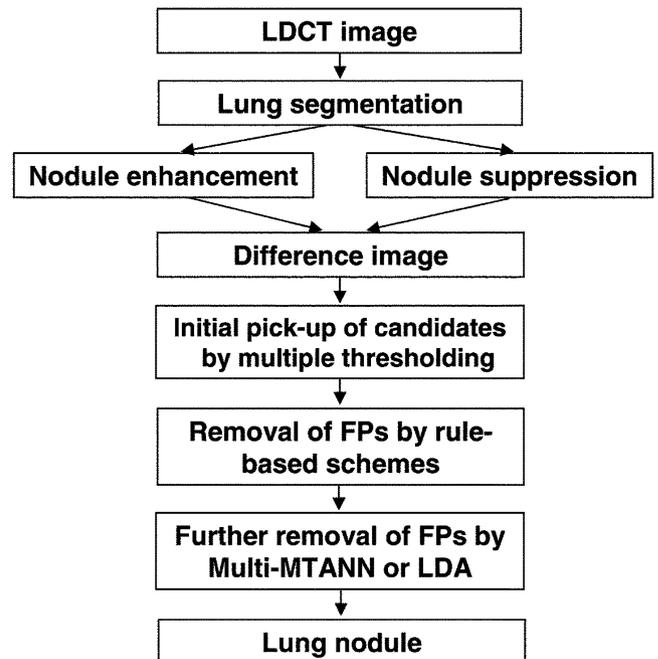


Figure 2. Overall scheme for computerized detection of pulmonary nodules on LDCT images based on a difference-image technique.

for dividing the histogram of the entire body region, lung regions were then segmented. The morphologic operation (closing operation) was applied for smoothing the outline of the segmented lung, and an image restoration technique (20) by use of the lung outline (filling in the entire region within the outline) was applied sequentially to the initially segmented lung for filling in holes in the segmented lung. If the area of the segmented lung was smaller than 450 mm², which was determined empirically, the slices with the smaller lungs were not applied for the subsequent processing. Each CT slice image was processed by two different filtering operations, namely, one for enhancement of nodules by use of a matched filter (8 mm nodule shape) and the other for suppression of nodules by use of a ring average filter (13 mm outside diameter, 0.6 mm width) (17). The difference image was then obtained by subtracting the nodule-suppressed image from the nodule-enhanced image. The effect of using the difference-image technique is shown in Fig 3, which depicts an original CT image with a missed cancer (detection error) overlapped with pulmonary vessels, and the corresponding difference image, where the cancer was well enhanced and small vessels were suppressed. The initial nodule candidates were selected on the difference image by use of a multiple gray-level thresholding technique.

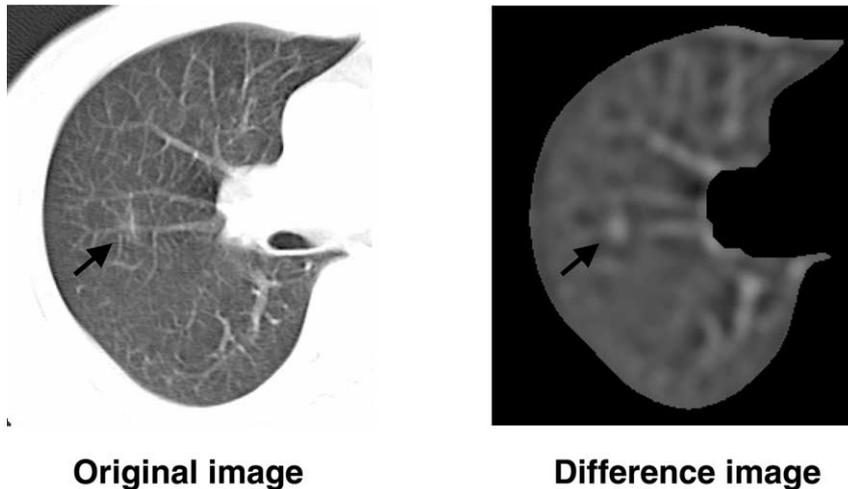


Figure 3. An original CT image with a missed cancer partially overlapped with small parallel vessel, and the corresponding difference image.

Next, regions of the nodule candidates were determined by monitoring the morphologic features with a region-growing technique. The false-positives among the candidates were reduced by applying the two rule-based schemes, first on the entire lung regions and second on divided lung regions. Finally, to further remove a number of false-positives, we separately applied two different classifiers, ie, a pattern-classification technique based on an artificial neural network (ANN), which is called a massive training artificial neural network (MTANN) (21–24), and LDA. By comparing the free response receiver operating characteristic (FROC) curves obtained by the two classifiers, we examined the usefulness of these classifiers.

Identification of Initial Nodule Candidates

Multiple gray-level thresholding was applied to the difference image for identification of initial nodule candidates (17). The pixel values above a given threshold level correspond with specific upper percentage of the area under the histogram. Each threshold level was determined according to a certain specific percentage of the area under the pixel-value histogram in the difference image from the high end of the histogram. Usually, the pixel values of nodules in the difference image are located at the high end of its histogram (the portion of the histogram closest to the highest pixel value). Therefore, we selected, empirically, specific percentages of 3% and 27% as the beginning and ending percentage threshold levels, respectively, with an increment of 3%. The regions in the difference image above a certain threshold value were

called “islands.” Islands with effective diameters smaller than 3 mm were excluded because the effective diameters of all cancers used in this study were larger than 6 mm, and the number of cancers less than 6 mm, which can be identified on LDCT images as “true” cancers, is generally very small (25), as will be addressed in the Discussion section. At the first percentage threshold level where each island emerged (referred to as “starting percentage threshold level”), the effective diameter and degree of circularity for the island were determined for selection of initial nodule candidates, because most of the nodules in the difference image are enhanced as being round, and the enhanced non-nodules (such as large vessels parallel to the cross-section) are not round and are larger than the nodules. A candidate selected at a starting percentage threshold level would not be examined again at the subsequent percentage threshold levels. The effective diameter of a candidate was defined by the diameter of a circle with the same area as that of the candidate. The degree of circularity was defined by the fraction of the overlap area of the circle with the candidate area. Note that nodules with various sizes and circularities tend to be picked up at low percentage threshold levels, whereas nodules with small size and large circularity tend to be picked up at high percentage threshold levels, as will be shown later. Thus, we designed the initial rule for picking up the nodule candidate at each starting percentage threshold level, for removal of some false-positives, if the circularity was lower than a certain threshold, and/or the size was larger than a certain threshold, as will be illustrated later. If the feature values of the island satisfied the initial rule for

picking up the nodule candidate at one of the starting percentage threshold levels, the island was considered an initial nodule candidate. Ten one-feature and one two-feature rules were determined for the initial pick-up of candidates.

Determination of Candidate Regions

To obtain the image features of the candidates for subsequent rule-based schemes, the candidate regions were determined by applying the region growing technique to the difference image and monitoring the morphologic image features (17). The region growing started at the location where the pixel value was a maximum in the area of the candidate, and was performed at various gray levels, which were decreased from each previous gray level with a decrement of a pixel value of 5 for low-contrast candidates or a pixel value of 10 for high-contrast candidates. The contrast was estimated for the purpose of determining a high or low contrast from the maximum pixel value in the candidate region on the initial identification of a nodule candidate (ie, whether the maximum pixel value was higher or lower than a specific pixel value), because pixel values on the background in the difference image were almost zero, and the maximum pixel value was very close to the contrast. At each gray level, the grown region of the candidate was quantified by the morphologic image features, ie, the effective diameter, circularity, and irregularity. The degree of irregularity was defined by $1-(P/N)$, where P is the perimeter of the circle and N is the length of the candidate outline. At a certain gray level, the effective diameters or the irregularity of many candidates tended to increase rapidly, but the circularity decreased steeply. This abrupt change in the size and shape of the candidate indicated that the candidate island merged with its surrounding background structures after that gray level. Therefore, at this transition point, the candidate regions and morphologic image features were determined. The transition point was defined at the gray level when the circularity decreased by more than 0.15 or when the irregularity or effective diameter increased by more than 0.20 or 5 mm, respectively. If the transition point was not detected, the candidate regions were determined when the circularity decreased below 0.6 or when the effective diameter increased above 10 mm. For the candidate for which the change did not occur, the candidate regions were determined at the last grown region, which was determined by region growing at a threshold value of 200.

Removal of False Positives by Two Rule-based Schemes

Two rule-based schemes were applied for removal of a number of false-positives first in the entire lung regions and second in the divided lung regions (inside and outside regions). In the first rule-based scheme for the entire lung regions, we determined the contrasts of the candidates on the difference image and original image, the standard deviation (SD) and contrast of pixel values in the outer region of candidate on the original image as well as the morphologic features, ie, the effective diameter, circularity, and irregularity of the candidate region as described above. The outer region of the candidate, where the SD and contrast were calculated, was defined as the outer region with a width of 3 pixels obtained from the candidate region by use of dilation of the morphologic filter with a 5-point kernel (four neighbors with a point of interest). The contrast used in this study was defined as the difference between the maximum pixel value and the minimum pixel value within a specific region (eg, the segmented candidate region or the outer region of a candidate region). The SDs and contrasts in the outer regions for candidates such as branching points of parallel vessels can be greater than those for nodules. The maximum pixel value of the candidate was obtained as the average pixel value in an area of 3×3 pixels over the pixel with the maximum value of the candidate. For determining the contrast on the original image, the pixel with the maximum value was searched in an area of 11×11 pixels of the original image, centering the pixel with maximum value in the candidate on the difference image.

In the second rule-based scheme, respective rules based on localized image features obtained from nodule candidates were established in divided lung regions for removal of the remaining false-positives. Each lung was divided into an outside region and an inside region, as shown in Fig 4a, because we believe that the characteristics (ie, image features) of false-positives in the outside region would be different from those in the inside region. For example, in the outside region, the effective diameters of vessels tended to be smaller or larger than those of the nodules. Therefore, we can remove some false-positives by selecting each rule in each region. The outside region and inside region in the lung were obtained by applying erosion of a morphologic filter with a circle kernel (10 mm diameter) to the segmented lungs, where the width (10 mm) of the outside region was determined empirically by observation of the effective diameter of vessels on the CT images. To analyze the localized image features of

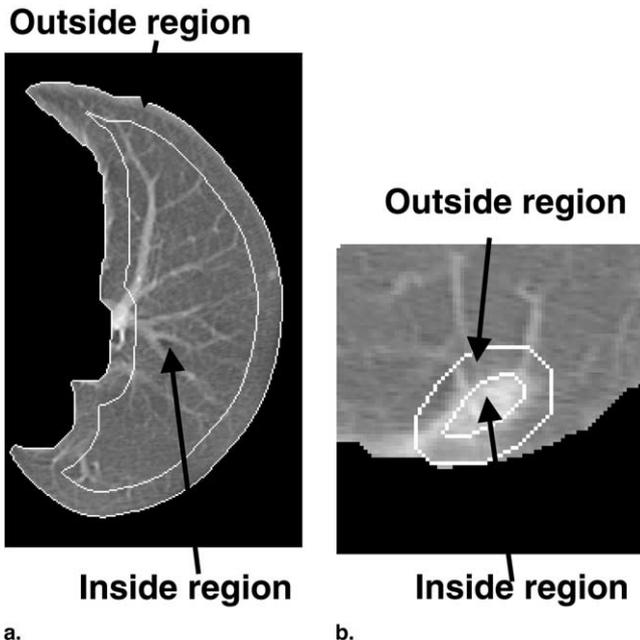


Figure 4. (a) Outside and inside regions in a lung used for determination of localized image features; (b) outside and inside regions of nodule candidate for determination of localized image features.

candidates, inside and outside regions for the candidate were obtained as shown in Fig 4b. The width of the outside region of the candidate obtained by use of dilation of the morphologic filter was 5 mm, where the appearance of nodules was different from that of vessels, especially parallel vessels. Because the pixel value distributions in the inside or outside region for some false-positives such as parallel vessels were different from those of the nodules, such false-positives could be distinguished from the nodules. Localized features calculated for the inside and outside regions of the nodule candidate were the average pixel value, full width at half maximum of the gray-level histogram, full width at tenth maximum of the gray-level histogram for the inside and outside regions, and the overlap area of gray-level histograms between the inside and outside regions, referred to as an overlap measure (26). In addition, the cross-correlation value between the difference image and the original image was obtained for eliminating the false-positives whose morphologic appearance on the difference image changed from that on the original image; however, the morphologic appearance of most nodules did not change appreciably. All of these features except the cross-correlation value were calculated in the original images.

False-positive Removal by use of Multi-MTANN or LDA Classifier

To reduce the false-positives of the nodule candidates in the final stage of the CAD scheme, the LDA classifier has commonly been used (11,13). Recently, a pattern-classification technique based on an ANN, the MTANN (21–24), was developed for removal of various types of false-positives produced by a CAD scheme developed by Armato et al (11). In our scheme, as a classifier, we applied the MTANN and the LDA separately, and we evaluated the overall performance shown by use of FROC curves, which were determined by changing the threshold for the multi-MTANN score or the LDA score. Finally, we examined the FROC curves obtained by MTANN and LDA, and the sensitivities at the same number of false-positives.

The MTANN consists of a modified multilayer ANN, which is capable of operating directly on the original image. The MTANN was 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 was obtained by scanning an input image with the MTANN. The distinction between a nodule and a non-nodule was made by use of a score that was defined from the output image of the trained MTANN. The multi-MTANN for eliminating various types of non-nodules consisted of plural MTANNs that were arranged in parallel. Each MTANN was trained by use of the same nodules but with a different type of false-positives, such as various sized vessels, and acted as an expert to distinguish nodules from a specific type of false-positive. The outputs of the MTANNs were combined by use of the logical AND operation so that each of the trained MTANNs did not eliminate any nodules, but removed some of the various types of false-positives. In this study, 15 MTANNs were configured, and the multi-MTANN was trained with 10 nodules (true-positives) obtained from nine scans (nine patients) and 150 false-positives from 51 scans (45 patients), which were produced by the second rule-based scheme with the missed cancer cases. All cases, including the training cases, were used for determining the FROC curve which indicates the overall performance of our CAD scheme using the multi-MTANN. Note, however, that the FROC curve obtained without the training cases was very similar to that obtained with all cases, where the sensitivities at 0.28 false-positives on two FROC curves were almost the same, as will be shown later in the Discussion section.

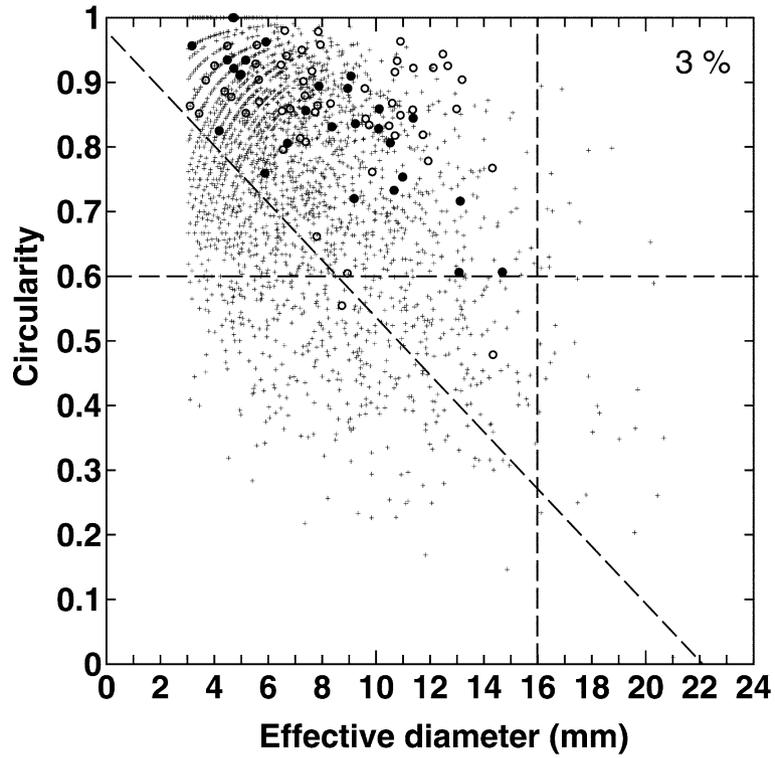
In our scheme, 18 image features can be used for the LDA, but some of the features were insignificant. To select the most effective image feature combination for separating the remaining candidates into the true-nodules and false-positives by use of a linear discriminant function, we determined the ROC curve for distinction between nodules and false-positives by using all data of true-positives and false-positives, and also used a stepwise method based on Wilks' lambda, which is defined by the ratio of within-group variance to the total variance (27), and the F value, which is a cost function based on Wilks' lambda. In the stepwise method, each feature was added or removed one-by-one by use of two thresholds on the F value, one for removal and another for addition (28), and the Az value, ie, the area under ROC curve, for each combination was computed for selection of the most effective feature combination with the highest Az value. Consequently, the final combination consisted of a starting percentage threshold level, effective diameter on the difference image, circularity on the difference image, SD of pixel values in the outer region of the candidate on the original image, overlap measure, average pixel value in the inside region, and full width at tenth maximum in the inside region. For determining the FROC curve of the CAD scheme using the LDA, a round-robin method per-patient basis (or leave-one-out-by-patient) was used. With this method, all candidates except for those obtained from the same patient were used for training, and candidates left out were used for testing the linear discriminant function. This procedure was repeated for all candidates, so that each candidate was used once as a test candidate.

RESULTS

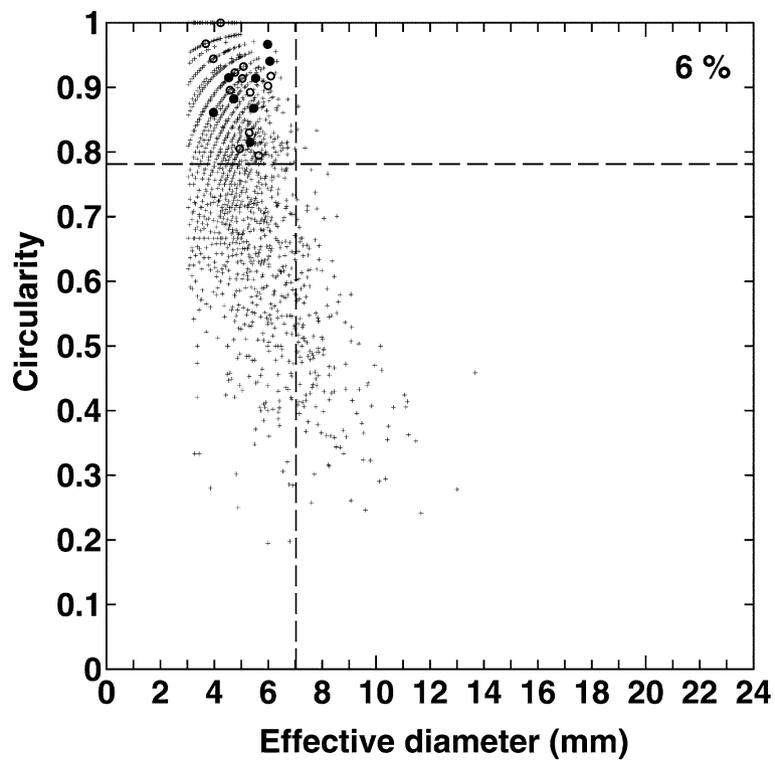
Our computerized scheme for nodule detection in LDCT images was applied to a cancer database of 106 scans (total number of slices, 3,292) with 131 nodules including 71 detected cancers and 38 missed cancers. As a result of lung segmentation, 524 of the 3,292 slices with small lung areas were excluded before use of the difference-image technique. Therefore, the number of CT images calculated for nodule detection processing in this database was 2,768, where the numbers of slices with and without nodules were 226 and 2,542. As shown in Figure 3, the missed cancer on the difference image was enhanced compared with that on the original image, and the small vessels adjacent to the nodule and background noise

were suppressed. However, some of the large parallel vessels, vertical vessels, and lung edge regions were also enhanced, and were included as false-positives in the initial nodule candidates. Because the difference-image technique was basically equivalent to a band-pass filter for enhancing the nodule-like objects, the vessels and lung edges were also enhanced with the filter.

Figures 5a and 5b demonstrate the initial pick-up rules in the relationship between the effective diameter and the circularity at starting threshold levels of 3% and 6%, respectively. These data for nodules and false-positives were obtained for determination of the initial pick-up rules from CT slices with 131 nodules by eliminating the candidates with the effective diameters smaller than 3 mm. The minimum circularity rule for the nodules increased from 0.6 to 0.8 with an increase in the percentage threshold level. At a 15% threshold level, the minimum circularity rule for the nodules was 0.9. On the other hand, the circularities of false-positives were frequently lower than those of nodules at each percentage threshold level, and the effective diameters of some false-positives were larger than those of the nodules. Therefore, by using the cut-off rules of the circularity and the effective diameter by the dashed lines shown in Figure 5, we removed a number of false-positives at each percentage threshold level and picked up initial nodule candidates. Because, as a result, the number of nodules which emerged at high percentage threshold levels of 18%, 21%, 24%, and 27% was only one, all nodule candidates at these levels were removed, and we examined nodule candidates only at percentage threshold levels of 3%, 6%, 9%, 12%, and 15%. After the initial pick-up of candidates, the two rule-based schemes were applied for removal of false-positives among the initial nodule candidates. Figure 6 shows one of the rules used in the first rule-based scheme in the relationship between the contrasts of nodule candidates on the original image and the difference image. Although "large" parallel vessels, vertical vessels, and lung edge regions were also enhanced, the contrasts of some false-positives such as "small" parallel vessels were suppressed on the difference image. Consequently, the difference between the contrasts of such false-positives in the original image and the difference image was greater than that of nodules. Thus, many false-positives were removed by the rule, as shown by the dashed line in Figure 6. Figure 7 shows one of rules used in the second rule-based scheme in the relationship between the effective diameter and the overlap measure, which is the overlap area between histo-



a.



b.

Figure 5. Relationship between the effective diameter and the circularity of nodules and false-positives (detected in the 127 slices with 131 nodules) at threshold levels of 3% (a) and 6% (b). Closed and open circles represent missed cancers and detected cancers, respectively, whereas small pluses represent false-positives. Rules are indicated by dashed lines.

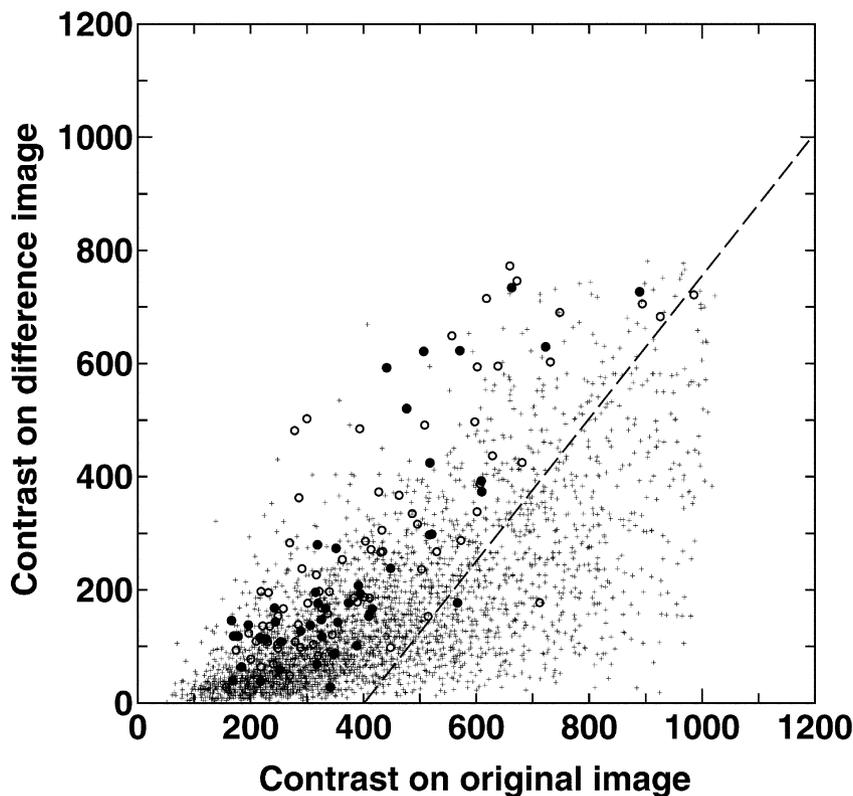


Figure 6. Relationship between the contrasts of nodule candidates (detected in the 127 slices with 131 nodules) on the original image and the difference image. Closed and open circles represent missed cancers and detected cancers, respectively, whereas small pluses represent false-positives. Rule is indicated by dashed line.

grams of the inside and outside regions for nodule candidates. These data were obtained in the inside regions of the divided lung regions. At each effective diameter, the overlap measures of some false-positives were greater than those of nodules. In the case of the candidates such as the branching point of parallel vessels, because the vessels extended from the inside region of the candidate to the outside region, the pixel-value histograms of the inside region for the branching point candidates were largely overlapped with those of the outside region compared with those of the nodules.

The sensitivity for all nodules and the number of false-positives per scan in the computerized detection of lung nodules at various steps are summarized in Table 1. The detection sensitivity for all nodules decreased from 93% to 81% with the three steps of false-positive removal, whereas the number of false-positives per scan was greatly decreased by each step. In the two rule-based schemes for false-positive removal, the number of false-positives per scan for all nodules decreased by 92% (316/343). Figure 8 shows the FROC curves for the overall

performance of our scheme by use of multi-MTANN or LDA. The sensitivity with LDA gradually decreased with a decrease in the number of false-positives, whereas the sensitivity with multi-MTANN remained constant until the number of false-positives per slice was reduced to 0.22 for all cancers or 0.28 for all nodules. Our scheme, by use of multi-MTANN or LDA, provided a sensitivity of 81% or 67% for all nodules, respectively, with 0.28 false-positives per slice. Therefore, multi-MTANN would be more appropriate for false-positive removal than LDA. Finally, 73% (19.7/27.0) of the remaining false-positives per scan were removed by using the multi-MTANN. As a result, our CAD scheme achieved a sensitivity of 83% (91/109) for all cancers with 5.8 false-positives per scan, and 81% (106/131) for all nodules with 7.3 false-positives per scan. The 84% (32/38) of missed cancers and 83% (59/71) of detected cancers were detected correctly with 5.9 and 5.8 false-positives per scan, respectively. Furthermore, it may be important to note that our CAD scheme detected 17 (74%) of 23 missed cancers, which were not mentioned in radiologists' clinical reports.

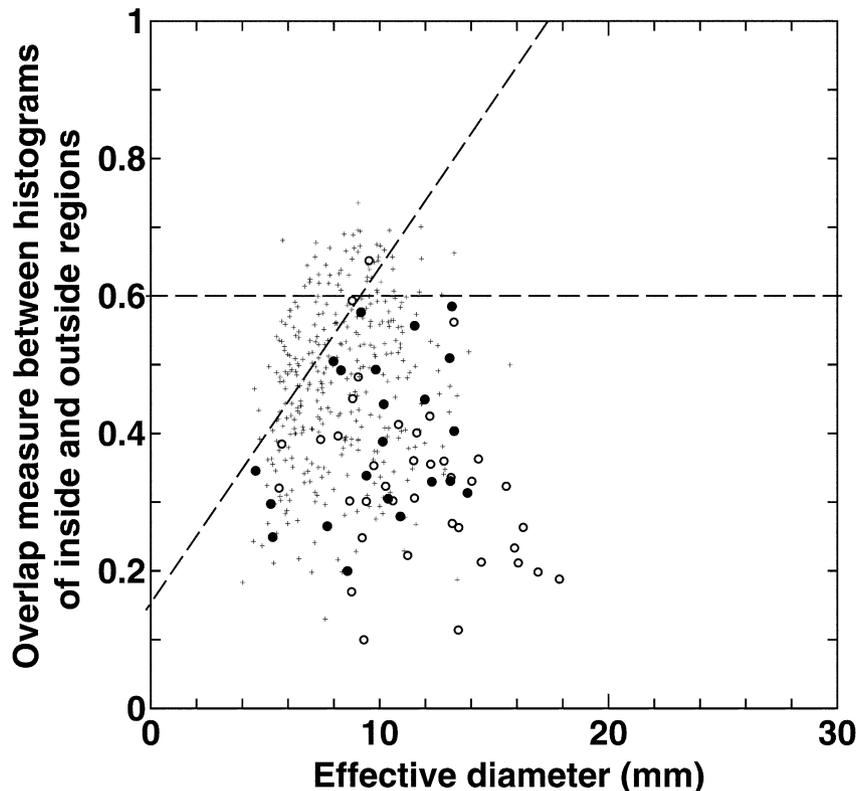


Figure 7. Relationship between the effective diameter and the overlap measure between histograms of inside and outside regions of nodule candidates (detected in the 127 slices with 131 nodules) in the inside regions of the lung regions. Closed and open circles represent missed cancers and detected cancers, respectively, whereas small pluses represent false-positives. Rules are indicated by lines.

DISCUSSION

Recently, Suzuki et al (23) applied the MTANN to the results obtained from a computerized scheme by Armato et al (11) by use of a database of 63 LDCT scans (a part of our database), where a gray-level thresholding tech-

nique was applied to CT images. The number of false-positives per scan was improved from 27.4 to 4.8 at a sensitivity of 80.3%. However, the difference between our method in this study and the method by Armato et al (11) was that the difference image technique by our method can enhance the nodules relative to normal background including small vessels, and thus the overall performance with our method without the multi-MTANN has been improved. For clarification, Table 2 shows the comparison of the performances together with databases used in this study and previous studies (11,23). According to Table 2, for the same database of 50 nodules with 38 “missed” cancers, our method achieved a sensitivity of 88% with 0.61 false-positives per slice before the multi-MTANN, and 88% with 1.9 false-positives per scan after the multi-MTANN, whereas the result by Armato et al (11) indicated a sensitivity of 80% with 1.0 false-positives per slice before the multi-MTANN, and 80% with 2.2 false-positives per scan after the multi-MTANN (23).

Table 1
Sensitivity and Number of False-Positives in Computerized Detection of Lung Nodules at Various Steps for 131 Nodules

	Sensitivity	No. of FPs Per Slice	No. of FPs Per Scan
Initial pick-up	93%	13.1	343
First rule-based scheme	85%	3.9	103
Second rule-based scheme	81%	1.0	27
Multi-MTANN	81%	0.28	7.3
LDA	67%	0.28	7.3

Note.—FPs = false-positives.

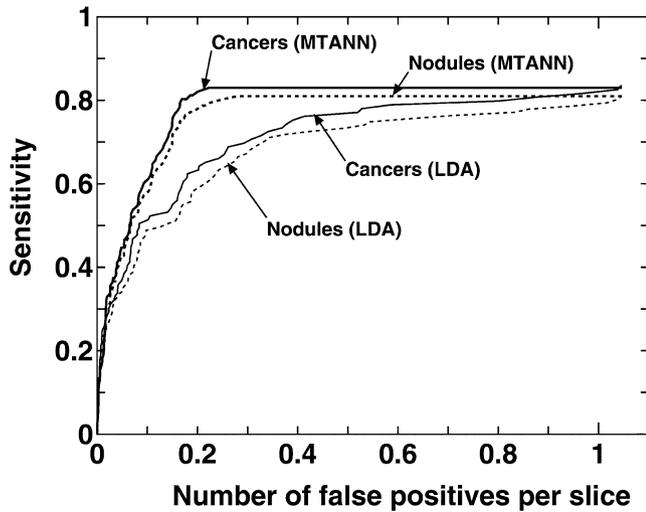


Figure 8. FROC curves of our CAD scheme for two databases (all cancers and all nodules) by use of multi-MTANN or LDA. All candidates in 2,768 slices were used for determination of this FROC curve.

The performance of multi-MTANN was superior to that of LDA, probably because multi-MTANN can learn various image features related to morphology and gray level directly from the images of typical nodules and false-positives, and

works as an image-based ANN filter with nodule enhancement and false-positive suppression to distinguish between nodules and false-positives. Furthermore, the multi-MTANN is a robust classifier for unknown cases. Figure 9 shows the comparison of FROC curves of our CAD scheme with the multi-MTANN obtained by use of the database with and without training cases. The FROC curve for the database without the training cases was similar to that for all cases, where the sensitivities at 0.28 false-positives on two FROC curves were almost the same. However, the use of multi-MTANN requires expertise for selection of typical nodules and typical false-positives and a considerable central processing unit (CPU) time for training (29.8 hours for a MTANN on a CPU; Pentium IV, 1.7 GHz), although the CPU time for testing was negligibly small. On the other hand, because LDA is based on a limited number of image features related to morphology and gray level, the amount of information used for LDA would be less than that for multi-MTANN; nevertheless, LDA is a simple and easily available classifier for most researchers compared with multi-MTANN.

Based on the 2-dimensional image features alone, our CAD scheme achieved a sensitivity of 83% for 109 can-

Table 2
Comparison of the Performances Together with the Methods and Databases used in this Study and Previous Studies in Armato and Colleagues (11,23)

	Current Study with Missed Cancer Database	Previous Studies with Missed Cancer Database (References 11,23)	Current Results with a Large Database	Previous Results with a Database Smaller than our Current Database (Reference 23)
Test database	38 scans with 50 nodules (38 missed cancers and 12 other nodules)	38 scans with 50 nodules (38 missed cancers and 12 other nodules)	106 scans with 71 confirmed cancers and 38 missed cancers (131 nodules: 109 cancers and 22 other nodules)	63 scans with 66 confirmed cancers
Training cases for detection scheme	all cases	all cases	all cases	all cases
Method for detection	difference-image technique	gray-level thresholding	difference-image technique	gray-level thresholding
Sensitivity with number of FPs per slice before Multi-MTANN	88% with 0.61 FPs per slice	80% with 1.0 FPs per slice (Ref. (11))	81% with 1.0 FPs per slice	82% with 1.0 FPs per slice
Training cases for Multi-MTANN	10 nodules from 9 scans; 150 FPs from 51 scans in a large database	10 nodules; 90 FPs from missed cancer database	10 nodules from 9 scans; 150 FPs from 51 scans in a large database	10 nodules; 90 FPs from missed cancer database
Sensitivity with number of FPs per slice (scan) after Multi-MTANN	88% with 0.07 FPs per slice (1.9 FPs per scan)	80% with 0.08 FPs per slice (2.2 FPs per scan) (Ref. 23)	81% with 0.28 FPs per slice (7.3 FPs per scan)	80% with 0.18 FPs per slice (4.8 FPs per scan) (Ref. 23)

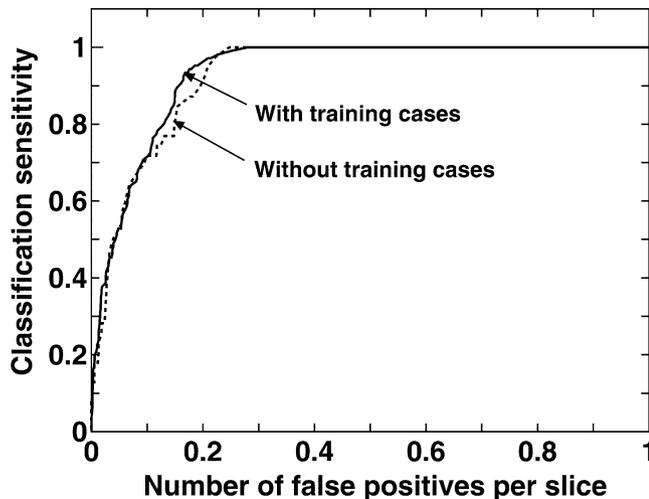


Figure 9. Comparison of FROC curves of our CAD scheme with the multi-MTANN obtained by use of the database with and without training cases.

cers with 5.8 false-positives per scan, which appears to be superior to results obtained in recent studies (11–13) in terms of the number of false-positives, where some false-positives such as lung vessels were removed by use of 3-dimensional image features. Therefore, by incorporating 3-dimensional image features in our scheme, the number of false-positives can be reduced further in the future.

Some of the nodules were not detected by our CAD scheme for two reasons. One was that some of the large and high-contrast nodules adjacent to the pleura were excluded during the initial segmentation process of the lungs. In the future, the initial segmentation process of the lungs should be improved so that such pleural nodules can be detected. It should be noted that all nodules with small size or low contrast were included in the segmented lungs obtained by use of our current scheme. Another reason was that some of the small and low-contrast nodules adjacent to or overlapped with the pleura or large vessels were not enhanced appreciably by use of the difference-image technique. For detecting such nodules, further study will be required for improving the filters used for enhancement of such nodules.

Recently, multidetector CTs with a thin slice thickness (eg, less than 2.5 mm or 1.0 mm) have been used for lung cancer screening, and many smaller nodules have been detected, compared with thick-slice CT. However, according to recent findings at the Mayo Clinic (25), 2,792 (98.6%) of 2,832 nodules detected by a multidetector CT were benign, 40 (1.4%) nodules were malignant, and four (0.14%) cancers were less than 6 mm. More-

over, only one (0.06%) of 1,735 nodules less than 4 mm was malignant. Therefore, we believe that, if our scheme based on the difference-image technique would be applied to such a database, small parallel vessels and background noise would be suppressed in the same way as that in thick-slice CT images, and the number of cancers “undetected” by our CAD scheme would be very small compared with all nodules “detected.” Thus, the clinical implication would be almost the same. Because a large number of small benign nodules detected by the multidetector CT is likely to produce many false-positives with a computerized detection scheme, whether a multidetector CT with thin slices should be used instead of thick-slice CT for early detection of lung cancer in screening programs is not obvious to us at present, and further investigation would be necessary.

In our scheme, 18 image features related to morphology and gray levels were determined, and a large number of physically understandable and logical rules were designed based on the relationships between image features. We believe that we were able to use a large number of rules in this study because we were able to define many additional features, because the difference image was created and because lung regions were divided into two areas, where rules were established in each region. For example, 60% of rules were based on two-feature rules in the relationship between an understandable image feature, ie, the effective diameter or the starting percentage threshold level (related to contrast) and the other image feature. We selected the useful relationship from many relationships for false-positive removal. We believe that it is important to present such new methods and/or new approaches to show the potential improvement in the performance of a CAD scheme. However, the generalization of a method is almost always a difficult issue in scientific papers. Although our study is based on a large screening program, it is still uncertain whether our scheme can perform at a level comparable to the result shown in this study when our scheme might be applied to another large screening program. Although some modifications would be necessary in the future for application to a general population, we believe that our scheme presented in this study would be useful as a foundation for further development. Nevertheless, we believe that a newly developed CAD system should be tested ideally with a large independent database that is different from the training cases. It should be noted that the performance of our CAD scheme derived from a consistency test may be overestimated, and the magnitude of the overestimate tends to be

larger with a complex classifier, as shown by Chan et al (29). However, it is very difficult at present to secure a large number of cancer cases for development and evaluation of CAD schemes for detection of lung nodules in CT images. This is one of the reasons that the Lung Image Database Consortium (30) has been established by the National Cancer Institute for creation of a large database that is available publicly. To our knowledge, our confirmed cancer database of LDCT acquired from a lung cancer screening is currently one of the largest databases available in this field. However, the number of cancers is still not adequate for an independent test. Thus, we recognize the need to test our CAD scheme in a clinical environment in the future.

In conclusion, we have developed a computerized scheme based on a difference-image technique for automated detection of lung nodules in LDCT images for lung cancer screening. By use of a database with 109 cancers including 38 missed cancers acquired from a lung cancer screening program, our CAD scheme achieved a sensitivity of 83% (91/109) of all cancers with 5.8 false-positives per scan. Furthermore, 84% (32/38) of all missed cancers were detected by our scheme with 5.9 false-positives per scan. Therefore, we believe that this computerized scheme would be useful for radiologists in detecting lung cancers on LDCT images for lung cancer screening.

ACKNOWLEDGMENT

The authors are grateful to Qiang Li, PhD, Hiroyuki Abe, MD, and Roger Engelmann, MS, for their useful discussions, and Elisabeth Lanzl for improving the manuscript.

REFERENCES

- Kaneko M, Eguchi K, Ohmatsu H, et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. *Radiology* 1996; 201:798-802.
- Sone S, Takashima S, Li F, et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. *Lancet* 1998; 351:1242-1245.
- Diederich S, Lenzen H, Windmann R, et al. Pulmonary nodules: experimental and clinical studies at low-dose CT. *Radiology* 1999; 213:289-298.
- Henschke CI, McCauley DI, Yankelevitz DF, et al. Early lung cancer action project: overall design and findings from baseline screening. *Lancet* 1999; 354:99-105.
- Nawa T, Nakagawa T, Kusano S, et al. Lung cancer screening using low-dose spiral CT: results of baseline and 1-year follow-up studies. *Chest* 2002; 122:15-20.
- Yamamoto S, Tanaka I, Senda M, et al. Image processing for computer-aided diagnosis of lung cancer by CT (LSCT). *Syst Comput Jpn* 1994; 25:67-79.
- Jiang H, Yamamoto S, Iisaku S, et al. Computer-aided diagnosis system for lung cancer screening by CT. In: Doi K, MacMahon H, Giger ML, Hoffmann KR, eds. *Computer-Aided Diagnosis in Medical Imaging*. Amsterdam: Elsevier, 1999; 125-130.
- Ukai Y, Niki N, Satoh H, et al. Computer aided diagnosis system for lung cancer based on retrospective helical CT image. *Proc SPIE (International Society for Optical Engineering)* 2000; 3979:1028-1039.
- Armato SG III, Giger ML, Moran CJ, et al. Computerized detection of pulmonary nodules on CT scans. *Radiographics* 1999; 19:1303-1311.
- Armato SG III, Giger ML, MacMahon H, et al. Automated detection of lung nodules in CT scans: preliminary results. *Med Phys* 2001; 28:1552-1561.
- Armato SG III, Li F, Giger ML, et al. Lung cancer: Performance of automated lung nodule detection applied to cancers missed in a CT screening program. *Radiology* 2002; 225:685-692.
- Wormanns D, Fiebich M, Saidi M, et al. Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system. *Eur Radiol* 2002; 12:1052-1057.
- Gurcan MN, Sahiner B, Petrick N, et al. Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. *Med Phys* 2002; 29:2552-2558.
- Brown MS, Goldin JG, Suh RD, et al. Lung micronodules: automated method for detection at thin-section CT-initial experience. *Radiology* 2003; 226:256-262.
- Giger ML, Doi K, MacMahon H, et al. Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields. *Med Phys* 1988; 15:158-166.
- Giger ML, Doi K, MacMahon H, et al. Pulmonary nodules: computer-aided detection in digital radiography. *Radiographics* 1990; 10:41-51.
- Xu XW, Doi K, Kobayashi T, et al. Development of an improved CAD scheme for automated detection of lung nodules in digital chest images. *Med Phys* 1997; 24:1395-1403.
- Li F, Sone S, Abe H, et al. Lung cancers missed at low-dose helical CT screening in a general population: comparison of clinical, histopathologic, and imaging findings. *Radiology* 2002; 225:673-683.
- Otsu N. A threshold selection method from gray level histograms. *IEEE Trans Syst Man Cybern* 1979; SMC-9:62-66.
- de Carmo MP. *Differential geometry of curves and surfaces*. Englewood Cliffs, NJ: Prentice Hall, 1976.
- Suzuki K, Armato SG III, Sone S, et al. Massive training artificial neural network for reduction of false positives in computerized detection of lung nodules in low-dose CT (Abstract). *Med Phys* 2002; 29:1322.
- Suzuki K, Armato SG III, Li F, et al. Computer-aided diagnostic scheme for detection of lung nodules in CT by use of massive training artificial neural network (Abstract). *Radiology* 2002; 225(P):533.
- Suzuki K, Armato SG III, Li F, et al. Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose CT. *Med Phys* 2003; 30:1602-1617.
- Suzuki K, Armato SG III, Li F, et al. 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. *SPIE (International Society for Optical Engineering) Proc* 2003; 5032:1355-1366.
- Swensen SJ, Jett JR, Hartman TE, et al. Lung cancer screening with CT: Mayo Clinic experience. *Radiology* 2003; 226:756-761.
- Aoyama M, Li Q, Katsuragawa S, et al. Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images. *Med Phys* 2002; 29:701-708.
- Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1992; Sec. 5.3:184-188.
- Sahiner B, Chan HP, Petrick N, Helvie MA, Goodsitt MM. Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis. *Med Phys* 1998; 24:516-526.
- Chan HP, Sahiner B, Wagner RF, et al. Effects of sample size on classifier design: quadratic and neural network classifiers. *Proc SPIE (International Society for Optical Engineering)* 1997; 3034:1102-1113.
- The Lung Image Database Consortium. Available at <http://www3.cancer.gov/dip/CII-PDF/Armato.pdf>