Pulmonary Nodules: Estimation of Malignancy at Thin-Section Helical CT—Effect of Computer-aided Diagnosis on Performance of Radiologists

**Kazuo Awai, MD**
**Kohei Murao, PhD**
**Akio Ozawa, BS**
**Yoshiharu Nakayama, MD**
**Takeshi Nakaura, MD**
**Duo Liu, MD**
**Koichi Kawanaka, MD**
**Yoshinori Funama, PhD**
**Shoji Morishita, MD**
**Yasuyuki Yamashita, MD**

**Purpose:**
To evaluate the effect of a computer-aided diagnosis (CAD) system on the diagnostic performance of radiologists for the estimation of the malignancy of pulmonary nodules on thin-section helical computed tomographic (CT) scans.

**Materials and Methods:**
The institutional review board approved use of the CT database; informed specific study-related consent was waived. The institutional review board approved participation of radiologists; informed consent was obtained from all observers. Thirty-three (18 malignant, 15 benign) pulmonary nodules of less than 3.0 cm in maximal diameter were evaluated. Receiver operating characteristic (ROC) analysis with a continuous rating scale was used to compare observer performance for the estimation of the likelihood of malignancy first without and then with the CAD system. The participants were 10 board-certified radiologists and nine radiology residents.

**Results:**
For all 19 participants, the mean area under the best-fit ROC curve ($A_z$) values achieved without and with the CAD system were $0.843 \pm 0.097$ (standard deviation) and $0.924 \pm 0.043$, respectively. The difference was significant ($P = .021$). The mean $A_z$ values achieved without and with the CAD system were $0.910 \pm 0.052$ and $0.944 \pm 0.040$, respectively, for the 10 board-certified radiologists ($P = .190$) and $0.768 \pm 0.078$ and $0.901 \pm 0.036$, respectively, for the nine radiology residents ($P = .009$).

**Conclusion:**
Use of the CAD system significantly ($P = .009$) improved the diagnostic performance of radiology residents for assessment of the malignancy of pulmonary nodules, however, it did not improve that of board-certified radiologists.

© RSNA, 2006

---

1 From the Department of Diagnostic Radiology, Graduate School of Medical Sciences (K.A., Y.N., T.N., D.L., K.K., S.M., Y.Y.), and Department of Radiological Technology, School of Health Sciences (Y.F.), Kumamoto University, 1-1-1 Honjiyo, Kumamoto 860-8556, Japan; and Bio-IT Business Development Group, Fujitsu, Chiba, Japan (K.M., A.O.). Received February 6, 2005; revision requested April 6; revision received May 10; final version accepted June 13. Address correspondence to K.A.

© RSNA, 2006
Screening for lung cancer by using low-radiation-dose helical computed tomography (CT) (low-dose CT) has gained attention in the past years. With low-dose CT, the detection rate has been observed to be 0.4%–2.7%, a value that is 2.6–10-fold higher than that for detection with chest radiography (1–10). Moreover, noncalcified nodules have been detected at 5%–66% of the examinations, a rate that is 1.7–3.0-fold higher than that for detection with chest radiography (1,2,6).

Sobue et al (11) reported that small lung cancers are associated with a better patient survival rate and that the 5-year survival rate was almost 100% in patients with nodules that were 9 mm or smaller. According to Henschke et al (6), in whose study 233 participants with one to six noncalcified nodules underwent low-dose CT, the largest nodule was 2–5 mm in 58%, 6–10 mm in 30%, 11–20 mm in 9%, and greater than 20 mm in 2% of the patients. Die- derich et al (7) also used low-dose CT for the detection of pulmonary nodules; in their study, the largest nodule was 2–5 mm in 73%, 6–10 mm in 23%, and greater than 10 mm in 3% of the 1001 participants. These findings suggest that low-dose CT is a promising method for the detection of small lung cancers.

Only 2.5%–11.6% of detected noncalcified nodules, however, prove to be lung cancer, and screening with low-dose CT results in many false-positive findings (3–7,9). Also, in a 1-year follow-up study with low-dose CT, noncalcified nodules were detected at 2.5%–3.9% of the examinations; only 3%–23% of these nodules were identified as lung cancer (1,5,10). Therefore, rational algorithms that facilitate the accurate diagnosis of noncalcified nodules detected at lung cancer screening with low-dose CT must be developed.

Thin-section CT has been recommended as the next step when a noncalcified nodule is detected at low-dose CT screening. At present, however, there are no clear diagnostic criteria for identification of malignant nodules detected by using thin-section CT, and in most instances, the interpretation of thin-section CT findings relies on the knowledge and experience of the radiologist who is performing the interpretation. The independent interpretation of noncalcified pulmonary nodules by two or more experienced radiologists and the use of a computer-aided diagnosis (CAD) system (12–36) for estimation of the malignancy of the nodules may assist radiologists in determination of a correct diagnosis. Thus, the purpose of our study was to evaluate the effect of a CAD system on the diagnostic performance of radiologists for the estimation of the malignancy of pulmonary nodules on thin-section helical CT scans.

Materials and Methods

All patients who underwent CT examinations at Kumamoto University Hospital, Kumamoto, Japan, gave prior informed consent for the use of their CT images in future retrospective studies. Our institutional review board approved the use of the CT database, and informed specific study-related consent was waived. Our institutional review board also approved the participation of radiologists in this observer performance study. Informed consent for the observer performance study was obtained from all observers.

Two authors (K.M. and A.O.) were employees of Fujitsu, and two other authors (K.A. and Y.Y.) had control of inclusion of all data and information for this study.

Nodule Selection

One chest radiologist (K.A.) with 18 years of chest CT experience reviewed the records of 171 consecutive patients who were suspected of having pulmonary nodules and underwent thin-section helical CT of the chest at our institute during a 24-month period from January 2002 to December 2003. He did not participate in the observer performance study. He selected all patients who satisfied the following criteria: They had only one pulmonary nodule that did not exceed 3 cm, and there was neither consolidation caused by the presence of organizing tissues after pneumonia nor consolidation associated with idiopathic pulmonary fibrosis around the nodules because it was difficult to define the boundary of the nodule. On the basis of these criteria, 21 malignant and 26 benign or possibly benign nodules were identified. Because three malignant nodules for which a histologic diagnosis was not determined were excluded, 18 malignant nodules were available for the observer performance test. Histologically, all 18 nodules were identified as primary lung cancer; 14 were well-differentiated adenocarcinomas and four were moderately differentiated adenocarcinomas. The diagnosis of these nodules was based on findings at video-assisted thoracic surgery in 14 nodules, findings at CT-guided transcutaneous biopsy in two nodules, and findings at bronchoscopic transbronchial biopsy in two nodules. We did not encounter any patients with solitary pulmonary metastasis during the period between January 2002 and December 2003.

Of the 26 possibly benign nodules, 15 were chosen for the observer performance test. In seven of these nodules, a

Abbreviations:
- AUC = area under the best-fit ROC curve
- CAD = computer-aided diagnosis
- ROC = receiver operating characteristic
- ROI = region of interest

Author contributions:
Guarantors of integrity of entire study, K.A., K.M., Y.Y.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, K.A., D.L.; clinical studies, A.O., Y.N., T.N., D.L., K.K., S.M., Y.Y.; statistical analysis, K.A., Y.F.; and manuscript editing, K.A.

See Materials and Methods for pertinent disclosures.
Estimation of Malignancy of Pulmonary Nodules at CT

Awai et al

Computerized Scheme for Differentiation of Pulmonary Nodules

First, one radiologist (K.A.) specified the region of interest (ROI) on a section that included the target nodule (Fig 1). The radiologist did not need to specify the ROI on the other sections because extension of the nodule along the z-axis was automatically estimated from the size of the ROI on the section. We determined the size of the ROI to encompass the nodule in all x-, y-, and z-axis directions and changed the size of the ROI in each nodule. For example, we used an ROI of 4–6 cm for the nodule that was 3 cm in maximal diameter. In the observer performance test described later, we set up the ROI for each case in advance and presented it on the screen of the cathode ray–tube monitor. Once the ROI was indicated, segmentation of the nodule was performed automatically with our CAD system.

The segmentation process was divided into three steps as follows:

1. Interpolation of image data was performed so that the pixel sizes in the x-, y-, and z-axis became isotropic. This preprocessing was necessary to segment the nodule with a smooth boundary and to quantify accurately the morphologic features of the pulmonary nodule.

2. Classification of the nodule type was then determined. Nodules were considered as solid, as a ground-glass opacity, or as that with a cavity. This classification was achieved by calculating the averaged CT number (Cav) of the small-volume area, which was a 1.5-mm cube around the center of the ROI. We calculated the Cav of the cube and defined the types as solid for Cav of −150 HU or greater, as ground-glass opacity for Cav of greater than −150 HU to −800 HU or greater, and as a nodule with a cavity for Cav of greater than −800 HU.

3. Elimination of structures tangent to the nodule, such as vessels and tho-

CT Imaging

CT imaging was performed with one of two scanners (LightSpeed QXi, GE Medical Systems, Milwaukee, Wis; Somatom Volume Zoom, Siemens Medical Systems, Erlangen, Germany). Scanning parameters for thin-section helical CT with the LightSpeed scanner were as follows: detector row width, 1.25 mm; helical pitch, 0.75; section thickness, 1.25 mm; section interval, 1.25 mm; rotation time, 0.8 second; tube voltage, 120 kVp; and tube current, 160–200 mA. For scanning with the Somatom Volume Zoom machine, the parameters were: detector row width, 1.0 mm; helical pitch (beam pitch), 0.75; section thickness, 1.0 mm; section interval, 1.0 mm; rotation time, 0.75 second; tube voltage, 120 kVp; and tube current, 160–200 mA.

Histologic diagnosis was established at video-assisted thoracic surgery. Among the histologically proved benign nodules, two were inflammatory lesions, two were focal fibrosis, one each was a granuloma and a pulmonary hamartoma, and one nodule was caused by anthracosilicosis. The other eight nodules either were histologically undetermined but manifested no change in size and internal and marginal characteristics at chest CT performed in the course of more than 2 years (six nodules) or disappeared at chest CT within 3 months (two nodules). The other 11 of 26 possibly benign nodules also manifested no change in size and internal or marginal characteristics at chest CT; however, observation periods for these nodules were shorter than 2 years (average, 10.7 months; range, 4–20 months). Therefore, we did not include these 11 nodules in the observer performance test.

The mean size in the xy (or transverse) plane of the 18 malignant nodules was 20.1 mm ± 8.9 (standard deviation), and that of the 15 benign nodules was 17.4 mm ± 9.1. According to the results of the two-tailed Student t test, there was no significant difference in size between malignant nodules and benign nodules (P = .394). The 33 patients were 15 men and 18 women who were 25–79 years old (mean age, 61.8 years); the mean age of patients with malignant nodules was 62.9 years (range, 25–79 years), and that of patients with benign nodules was 60.6 years (range, 39–72 years). According to the results of the two-tailed Student t test, there was no significant difference between the age of patients with malignant nodules and that of patients with benign nodules (P = .381).

Figure 1: Example of an estimation of the likelihood of malignancy of a pulmonary nodule on a transverse thin-section CT image. In the main window, the target nodule is indicated by the octagon. The pathologic diagnosis was well-differentiated adenocarcinoma. In the enlarged window at right are the results of the estimation. The last row shows the estimation of likelihood of malignancy.
radic walls, was performed. This elimination was achieved by using several image processing techniques, such as “snakes” (37) and mathematical morphology (38). For example, vessels around the nodule were eliminated by using the “geodesic erosion/dilation” technique of mathematical morphology (38). At first, a lesion area within the ROI was extracted according to the threshold of the CT number. Then, the margin of the lesion area was eroded by a structure element that had almost the same diameter as the vessels adjacent to the nodule. Finally, the lesion area was dilated so as to reproduce the accurate margin of the lesion area without vessels.

In the next step, morphologic features of the nodule were quantified on the basis of the segmentation results with our CAD system (Table 1). We selected morphologic features that were effective for estimation of the malignancy of the nodule by investigating the strength of the neural network connection (39). In the training stage, a neural network (40,41) was established by using an iterated calculation to minimize the difference between pathologic results and the outputs from the morphologic features. Thirty-four nodules were used for the training of our CAD system, and these cases were not included among the cases used for the observer performance test.

In the differentiation stage, the morphologic features of undiagnosed nodules were fed to the neural networks. The output was the likelihood of malignancy that was represented by the continuous values ranging from 0% to 100%. At the observer performance test, we also showed the results of analysis of the morphologic features. The values of maximal diameter, volume, mean, and standard deviation of the CT number of the nodule were shown in a text style so that radiologists could check whether the segmentation results were acceptable in comparison with their visual observations. The values of features such as inner inhomogeneity of CT number, marginal obscurity, irregularity of the surface, and sphericity were shown in a bar chart. The height of the bar was normalized according to the statistics of the training data set for the CAD system. The value of the feature $F$ was normalized to $\tilde{F}$ according to the following equation:

$$\tilde{F} = \frac{F - F_{av}}{2F_{av}} + 50.$$  

Here, $F_{av}$ and $F_{\sigma}$ were the average and the standard deviation, respectively, of the morphologic features of the training set for use of the CAD system (Fig 2).

A computer workstation (Primergy; Fujitsu, Tokyo, Japan) with dual 1.4-GHz processors (Pentium III; Intel, Santa Clara, Calif) was used in this study, and the average calculation time for the CAD system to analyze each nodule was about 2.5 minutes.

### Observer Performance Study

With a sequential-test method, we employed receiver operating characteristic

### Table 1

<table>
<thead>
<tr>
<th>Morphologic Feature</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density</strong></td>
<td></td>
</tr>
<tr>
<td>Averaged CT number</td>
<td>None</td>
</tr>
<tr>
<td>Standard deviation of CT number</td>
<td>None</td>
</tr>
<tr>
<td>Internal inhomogeneity of CT number</td>
<td>Nonuniformity of gradient vectors inside solid part</td>
</tr>
<tr>
<td>Marginal obscurity</td>
<td>Volume rate of lower CT number part tangent to solid part</td>
</tr>
<tr>
<td><strong>Shape</strong></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td>None</td>
</tr>
<tr>
<td>Sphericity</td>
<td>None</td>
</tr>
<tr>
<td>Irregularity of surface</td>
<td>Nonuniformity of normal vectors on surface of nodule</td>
</tr>
<tr>
<td>Maximal diameter</td>
<td>None</td>
</tr>
<tr>
<td>Calcification volume*</td>
<td>Volume rate of calcification to volume of whole nodule</td>
</tr>
<tr>
<td>Cavity volume*</td>
<td>Volume rate of cavity to volume of whole nodule</td>
</tr>
</tbody>
</table>

* The value was calculated as a percentage.
(ROC) analysis to evaluate the performance of our CAD system and that of the 19 participating radiologists in distinguishing between benign and malignant nodules without and with the CAD output (42). The observers were 10 board-certified radiologists with 8–26 years of experience (mean, 17.0 years) and nine radiology residents with 1–4 years of experience (mean, 2.1 years). All board-certified radiologists specialized in body imaging and read thoracic CT images on a regular basis.

Two clinical parameters (patient age and sex) were disclosed to the observers on the cathode-ray tube monitor. They were allowed to change the level and width of the window on the monitor; reading time was not limited. The observers were not given specific criteria for judging a nodule as benign or malignant; rather, they were asked to use their existing knowledge. In response to the question of whether they judged the nodule to be benign or malignant, the radiologists marked their confidence level in regard to the likelihood of malignancy on a continuous rating scale (43,44). By using a mouse, they indicated their judgment on a horizontal bar displayed on the screen; “definitely malignant” and “definitely benign” were marked at the right and left ends of the bar, respectively.

CT images were first presented without the CAD output. After each radiologist marked the initial level of confidence, the computer output for the results of quantitative measurements and of the analysis for likelihood of malignancy were displayed on the monitor. Each observer then had a chance to change the previously indicated confidence level.

Before training and test taking, the participating radiologists were informed that the purpose of the experiment was to evaluate the potential benefit of using the CAD system to help in the differentiation between malignant and benign pulmonary nodules on thin-section helical CT scans. They were also told that 33 nodules would be shown randomly, that approximately 50% of the nodules were malignant, and that the accuracy of the computer output was about 73% when a threshold level of 50% for the estimated degree of malignancy with the computer was used as the measure of the likelihood of malignancy. They were instructed to use the rating scale consistently and uniformly.

All participants were trained with three training cases of pulmonary nodules before they participated in the observer performance test to make sure that they could operate the observer interface and they knew how to take into account the computer output in their decision making. Cases used for training were not included among those used for the observer performance test.

To assess the diagnostic difficulty level of each observer performance test case, we recorded the number of radiologists who, without the CAD system, incorrectly answered whether the nodule was malignant or benign in cases in which a threshold level of 50% was used as the measure of the likelihood of malignancy.

**Statistical Analysis**

ROC analysis was used to compare the radiologists’ performance without and with CAD output in distinguishing between benign and malignant pulmonary nodules. A binormal ROC curve was fitted to each radiologist’s confidence level in regard to the likelihood of malignancy. They were instructed to use the rating scale consistently and uniformly. All participants were trained with three training cases of pulmonary nodules before they participated in the observer performance test to make sure that they could operate the observer interface and they knew how to take into account the computer output in their decision making. Cases used for training were not included among those used for the observer performance test.

To assess the diagnostic difficulty level of each observer performance test case, we recorded the number of radiologists who, without the CAD system, incorrectly answered whether the nodule was malignant or benign in cases in which a threshold level of 50% was used as the measure of the likelihood of malignancy.

**Table 2**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Board-certified Radiologists</th>
<th>Radiology Residents</th>
<th>All Radiologists</th>
<th>CAD System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>76.1</td>
<td>67.3</td>
<td>71.9</td>
<td>72.2</td>
</tr>
<tr>
<td>Specificity</td>
<td>89.3</td>
<td>74.8</td>
<td>82.5</td>
<td>75.0</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>90.0</td>
<td>77.2</td>
<td>84.0</td>
<td>81.3</td>
</tr>
</tbody>
</table>

Note.—All numbers are percentages.

**Figure 3**

Graph shows number of radiologists who incorrectly identified 33 nodules as malignant or benign without CAD system.
with CAD output was tested by using the same computer program; differences were estimated by using analysis of variance for pseudovalues of $A_z$ calculated from the rating scores of all 19 participating radiologists.

The significance of the difference between $A_z$ values was determined. These differences were between the board-certified radiologists who performed assessments without or with the CAD system and the radiology residents who performed assessments with the CAD system. They were also between the board-certified radiologists who performed assessments without the CAD system and the radiology residents who performed assessments without the CAD system. Evaluation of these differences was performed with multiple comparisons by using the Tukey-Kramer method. Statistical analyses were performed with a statistical software package (SPSS, version 11.0; SPSS, Chicago, Ill). $P$ values of less than .05 were considered to indicate a significant difference.

We also calculated the sensitivity, specificity, and positive predictive values for the diagnosis of malignant nodules determined by the board-certified radiologists and by the radiology residents.

**Results**

Of the 18 malignant nodules, three were incorrectly diagnosed by more than three board-certified radiologists, and nine were incorrectly diagnosed by more than three radiology residents. Of the 15 benign nodules, five were incorrectly diagnosed by more than three board-certified radiologists, and seven were incorrectly diagnosed by more than three residents (Fig 3). When a threshold level of 50% was used for the measure of the likelihood of malignancy in regard to the confidence level, the sensitivity, specificity, and positive predictive values of our CAD system alone were 72.2%, 75.0%, and 81.3%, respectively (Table 2). All sensitivity, specificity, and positive predictive values for the CAD system were between the values for board-certified radiologists and those for the radiology residents (Table 2). The $A_z$ value for the CAD output alone was 0.795.

![Figure 4](image)

**Figure 4:** Mean ROC curves for all observers who distinguished between benign and malignant nodules without and with CAD output and ROC curve for the CAD output alone. The mean $A_z$ values for all radiologists increased from 0.843 ± 0.097 without CAD output to 0.924 ± 0.043 with CAD output. The difference was significant ($P = .021$). The $A_z$ value for the CAD output alone was 0.795.

### Table 3

$A_z$ Values for Board-certified Radiologists and Radiology Residents for Diagnosis without and with CAD Output

<table>
<thead>
<tr>
<th>Observer</th>
<th>$A_z$ Value without CAD Output</th>
<th>$A_z$ Value with CAD Output</th>
<th>$A_z$ Value with CAD Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Board-certified radiologists</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.871</td>
<td>0.890</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.808</td>
<td>0.888</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.871</td>
<td>0.939</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.883</td>
<td>0.896</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.927</td>
<td>0.953</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.939</td>
<td>0.981</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.953</td>
<td>0.974</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.929</td>
<td>0.980</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.940</td>
<td>0.947</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.984</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>Average*</td>
<td>0.910 ± 0.052</td>
<td>0.943 ± 0.040</td>
<td></td>
</tr>
<tr>
<td>Radiology residents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.836</td>
<td>0.931</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.811</td>
<td>0.874</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.776</td>
<td>0.885</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.761</td>
<td>0.874</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.666</td>
<td>0.902</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.754</td>
<td>0.883</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.653</td>
<td>0.935</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.749</td>
<td>0.860</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0.905</td>
<td>0.969</td>
<td></td>
</tr>
<tr>
<td>Average*</td>
<td>0.768 ± 0.078</td>
<td>0.901 ± 0.036</td>
<td></td>
</tr>
<tr>
<td>Average for all*</td>
<td>0.843 ± 0.097</td>
<td>0.924 ± 0.043</td>
<td></td>
</tr>
</tbody>
</table>

* Data are the mean ± standard deviation.
with CAD output were 0.768 ± 0.078 and 0.901 ± 0.036, respectively; these values were significantly different \( (P = 0.009) \) (Table 3).

There was a significant difference between the mean \( A_z \) values recorded for board-certified radiologists and those recorded for the residents when assessment was performed without the CAD system \( (P < 0.001) \). There was no significant difference, however, between the mean \( A_z \) values recorded for board-certified radiologists who did use the CAD system or those who did not use the CAD system, on one hand, and those for the group of residents who used the CAD system in their assessment, on the other \( (P = .892 \text{ and } P = .101, \text{ respectively}) \).

**Discussion**

The spread of lung cancer screening with low-dose CT has drawn attention to the importance of the use of a CAD system for the assessment of pulmonary nodules \((12-21,23-31,33-36,39,45,46)\). We found that, when radiology residents used the CAD system, their diagnostic abilities advanced to the same level as that of board-certified radiologists who performed the assessment with or without the CAD system. On the other hand, use of the CAD system did not significantly enhance the diagnostic abilities of board-certified radiologists. Of 18 malignant nodules, three (17%) were incorrectly diagnosed by more than three board-certified radiologists, as were five (33%) of the 15 benign nodules. On the other hand, nine (50%) malignant nodules and seven (47%) benign nodules were incorrectly diagnosed by more than three residents. This finding suggests that the data sets presented in our observer performance test may have been relatively easy to diagnose for board-certified radiologists and may have been relatively difficult for the radiology residents to diagnose. The \( A_z \) value for our CAD system was smaller than that for the board-certified radiologists as a group and larger than that for the residents as a group. To render our CAD system useful for not only less experienced residents but also experienced board-certified radiologists, its performance must be raised to the diagnostic ability level of board-certified radiologists.

Our CAD system displays not only estimations of malignancy in percentages but also typical parameters in numeric values or in histogram formats. This may make it possible for the observer to use the displays for assessment of the appropriateness of the malignancy estimations calculated with the CAD system. We did not, however, confirm that results of morphologic feature analysis actually contributed to final decisions of radiologists.

With our CAD system, estimates of the likelihood of malignancy that are based on morphologic features of the pulmonary nodules on thin-section helical CT scans were calculated. The addition of clinical information, such as patient age, sex, and smoking history, may improve the diagnostic performance of our system. Furthermore, although the neural network was used with our CAD system to estimate the likelihood of malignancy on the basis of various morphologic features, Bayesian analysis \((47-49)\) is a viable alternative.

With our CAD system, estimates of the likelihood of malignancy are determined by using learning input data. When the CAD system contains data on benign nodules, such as a pulmonary hamartoma, but not on metastatic pulmonary tumors, the latter may be estimated as benign. Therefore, it is important to recognize that the estimation of the likelihood of malignancy that is based on the assessment of only CT images is not totally reliable. Even when the CAD system returns a diagnosis of a benign pulmonary nodule, if this estimation is based on a single CT image, it is advisable to monitor the nodule for temporal changes. Our CAD system can be used to monitor temporal changes from various perspectives because the results of analysis that can be performed in regard to internal density and marginal characteristics, as well as volumetry, are considered.

We used a threshold level of 50% for the measure of the likelihood of malignancy with our CAD system. This threshold level was used to assess the sensitivity, specificity, and positive predictive values for our CAD system, for board-certified radiologists, and for radiology residents for the estimation of malignancy. Much lower probabilities of malignancy can and should result in a change in patient treatment. For example, some decision-analysis studies suggest that a probability of 5%–10% or
greater should result in a change in treatment (50,51). The likelihood of malignancy that was estimated by using the CAD system represents only a reference for probability of malignancy, and radiologists or pulmonologists must determine the appropriate treatment for pulmonary nodules.

We used our CAD system to perform quantitative analysis of pulmonary nodules displayed on thin-section helical CT scans. If thin-section helical CT is employed for lung cancer screening, the detection and characterization of tumors can be achieved simultaneously. However, small nodules can be identified (52), and the misidentification of structures, such as blood vessels, as nodules may cease. Thin-section helical CT, however, generates a large volume of image data and increases the amount of radiation exposure to patients.

There were some limitations in our study. The nodules that were investigated were almost 3 cm in diameter. We chose tumors that were less than 3 cm in diameter because this size corresponds with stage T1 tumors according to the Union Internationale Contre le Cancer classification. Patients whose nodules are 2 cm or larger and are without calcification usually undergo biopsy or surgery immediately (4), and most nodules that require CT for estimation of their benign or malignant nature are relatively small (2 cm in maximal diameter). We are preparing to conduct an observer performance test in which we include nodules with a diameter of only 2 cm or smaller. Also, only 33 lesions were presented in our observer performance test to allow individual observers to perform their interpretation in a relatively short time. On the other hand, 19 radiologists with different levels of experience participated in our study; this number of observers was sufficient large for validation of the reliability of our statistical results.

In conclusion, use of our CAD system significantly improved the diagnostic performance of radiology residents but not of board-certified radiologists. Our future challenge is to increase the performance level of the CAD system to a level identical to, or higher than, the performance level of the board-certified radiologists.

Acknowledgments: The authors are grateful to Lorenzo Pesce, PhD (University of Chicago, Department of Radiology, Kurt Rossmann Laboratories, Chicago, Ill), for providing the computer program (LABMRMC, version 1.483) for statistical analysis and for helpful advice about the statistics.

References

25. Lawler LP, Wood SA, Pannu HK, Fishman


