Computer-aided Diagnosis for the Detection and Classification of Lung Cancers on Chest Radiographs:

ROC Analysis of Radiologists' Performance[1]

Junji Shiraishi, PhD, Hiroyuki Abe, MD, PhD, Feng Li, MD, PhD, Roger Engelmann, MS
Heber MacMahon, MB, Kunio Doi, PhD

Rationale and Objectives. The aim of the study is to investigate the effect of a computer-aided diagnostic (CAD) scheme on radiologist performance in the detection of lung cancers on chest radiographs.

Materials and Methods. We combined two independent CAD schemes for the detection and classification of lung nodules into one new CAD scheme by use of a database of 150 chest images, including 108 cases with solitary pulmonary nodules and 42 cases without nodules. For the observer study, we selected 48 chest images, including 24 lung cancers, 12 benign nodules, and 12 cases without nodules, from the database to investigate radiologist performance in the detection of lung cancers. Nine radiologists participated in a receiver operating characteristic (ROC) study in which cases were interpreted first without and then with computer output, which indicated locations of possible lung nodules, together with a five-color scale illustrating the computer-estimated likelihood of malignancy of the detected nodules.

Results. Performance of the CAD scheme indicated that sensitivity in detecting lung nodules was 80.6%, with 1.2 false-positive results per image, and sensitivity and specificity for classification of nodules by use of the same database for training and testing the CAD scheme were 87.7% and 66.7%, respectively. Average area under the ROC curve value for detection of lung cancers improved significantly ($P = .008$) from without (0.724) to with CAD (0.778).

Conclusion. This type of CAD scheme, which includes two functions, namely detection and classification, can improve radiologist accuracy in the diagnosis of lung cancer.

Key Words. Computer-aided diagnosis (CAD); ROC; lung cancer; lung nodule

© AUR, 2006

Because some evidence suggests that early detection of lung cancer may allow timely therapeutic intervention and thus a more favorable prognosis for the patient (1–5), lung cancer screening by use of chest radiographs and low-dose computed tomographic (CT) scans has been proposed (6–8), although this is a highly controversial issue (9) at present. Chest radiography, which is simple to perform and inexpensive, commonly has been used as a first examination for detecting lung cancers; however, it was well shown that detecting lung cancer at an early stage on chest radiographs is a very difficult task for radiologists (6,10). Although CT is more sensitive than chest radiography in the detection of small noncalcified nodules (8), it is very difficult to distinguish benign from malignant lesions, and interpretation of a large number of CT images is time consuming for radiologists. Therefore, a number

Acad Radiol 2006; 13:995–1003

[1] From the Department of Radiology, Kurt Rossmann Laboratories for Radiologic Image Research, The University of Chicago, 5841 South Maryland Avenue, MC2026 Chicago, IL 60637. Received February 2, 2006; revision received April 17; revision accepted April 19. **Address correspondence to:** J.S. e-mail: junji@uchicago.edu

© AUR, 2006
doi:10.1016/j.acra.2006.04.007

of investigators are developing computer-aided diagnostic (CAD) schemes to facilitate early identification of lung cancer, including the detection of lung nodules on chest radiographs (11–15) and CT images (16–21), and the distinction between benign and malignant lung nodules on chest radiographs (22,23) and CT images (24–26).

A number of CAD schemes were developed for the detection of lung nodules and also for distinction between benign and malignant lung nodules in chest radiographs during the last 20 years (11–15,22,23), and several observer performance studies were carried out to show the usefulness of the application of each CAD scheme for radiologist decision making (22,27–30). However, in terms of lung cancer diagnosis on chest radiographs, the effect of these CAD schemes on radiologist decision making may be more complicated because radiologists are likely to perform two tasks simultaneously for detection and classification. To introduce these computerized schemes into clinical practice, we believe the development and evaluation of an integrated computerized scheme for such multiple functions as detection and classification is necessary.

In this study, we combine two independent CAD schemes for the detection and classification of lung nodules to evaluate the performance of radiologist decision making for lung cancer diagnosis on chest radiographs when the CAD scheme provided radiologists with both the location and likelihood of malignancy of suspected lung nodules.

MATERIALS AND METHODS

Our institutional review board approved the use of the image database and participation of radiologists in this observer performance study. Informed consent for each case was waived because this is a retrospective study, and informed consent for the observer performance study was obtained from all observers.

Image Database

We used 150 chest images, which consisted of 108 cases with solitary lung nodules (75 malignant and 33 benign nodules) and 42 non-nodule cases, for training and testing the computerized schemes. For the observer study, we selected 48 cases (24 malignant, 12 benign, and 12 non-nodule cases) from the total of 150 cases by using a stratified sampling method (31), which can keep the fraction of nodules constant in terms of their degrees of subtlety.

All chest images used in this study were selected from the Japanese Standard Digital Image (JSRT) Database developed by the Japanese Society of Radiological Technology (32), which is available publicly. The JSRT Database includes 154 abnormal chest radiographs, each with a solitary pulmonary nodule, and 93 non-nodule chest radiographs. These original screen-film images were digitized with a 0.175-mm pixel size, matrix size of 2048 × 2048, and 12 bits of gray scale. All cases in the JSRT Database were confirmed by means of CT examination regarding the presence or absence of a lung nodule. All nodule cases were classified as malignant based on histologic and cytologic examination or as benign based on histologic examination, definitive isolation of a pathogenic organism, shrinkage and disappearance with the use of antibiotics, or no change observed during a 2-year follow-up. Nodule images included in the JSRT Database were divided into five subjective rating groups according to the degree of subtlety of the lung nodule. Numbers of cases in the database for each group according to degree of subtlety were 25 extremely subtle cases, 29 very subtle cases, 50 subtle cases, 38 relatively obvious cases, and 12 obvious cases. Degree of subtlety correlated closely with level of radiologist performance in the detection of lung nodules, which was shown in receiver operating characteristic (ROC) analysis by 20 radiologists (32).

We selected 108 lung nodule cases from the original 154 cases, with several exclusion criteria determined by two chest radiologists (H.A. and F.L.) as follows: (1) nodules categorized as either extremely subtle or extremely obvious (n = 37), (2) nodule size larger than 35 mm on the chest radiograph (n = 3), (3) another unconfirmed abnormality (ie, nodule) in the image (n = 5), and (4) contour of nodule not well defined (n = 1). In addition, 42 non-nodule cases were selected randomly from the original 93 non-nodule cases. Table 1 lists final diagnoses for the 108 nodule cases (75 malignant and 33 benign) and selected 36 cases (24 malignant and 12 benign) used in the observer performance study. Mean sizes of the 75 malignant and 33 benign nodules were 17.8 mm (range, 9.5–31.1 mm) and 15.2 mm (range, 8.9–28.8 mm), respectively.

Computerized Scheme for Detection of Lung Nodules

The computerized nodule detection scheme used in this study consisted of five steps, ie, (1) preprocessing and segmentation of lung fields by use of ribcage edge detection techniques (33,34), (2) identification of initial nodule

Table 1
Diagnoses of 108 Nodule Cases Used in This Study for Training and Testing the CAD Scheme and 36 Nodule Cases for the Observer Performance Study

	Cases Selected for CAD Development (108)	Cases Selected for ROC Study (36)
Malignant nodules	75	24
Adenocarcinoma	19	6
Bronchioloalveolar cell carcinoma	6	1
Small-cell carcinoma	1	0
Large-cell carcinoma	1	0
Adenosquamous carcinoma	3	0
Lung cancer (unknown subtype)	41	15
Metastasis	4	2
Benign nodules	33	12
Inflammatory lesion	12	3
Tuberculoma	9	3
Granuloma	6	2
Hamartoma	3	2
Unknown	3	2

candidates by use of a multiple-thresholding technique (13) with a difference image obtained from the subtraction of signal-enhanced and signal-suppressed images (11), (3) extraction of image features from original and difference images (13), (4) elimination of false-positive candidates by use of rule-based tests (13), and (5) reduction of remaining false-positive results by use of linear discriminant analysis (LDA) (35).

The performance level used in this study for the detection of lung nodules indicates that sensitivity was 80.6%, with 1.2 false-positive results per image, when the 108 nodule and 42 non-nodule cases were used for training and testing the computerized scheme.

Computerized Scheme for Classification of Lung Nodule Candidates

The computerized scheme is designed for distinguishing malignant from benign nodules, as well as computer-detected and radiologist-detected false-positive candidates, the locations of which were identified by the computer and/or radiologists on digital chest images. Our scheme consisted of three steps, ie, (1) automated segmentation at the identified location by analysis of contour lines of the gray-level distribution based on a polar-coordinate representation (23), (2) extraction of image features determined from the outline and image analysis for regions inside and outside the segmented nodule, and (3) LDA for determining the likelihood of malignancy of each nodule candidate. Seven features used in LDA included patient age, the root-mean-square value of the power spectrum, an overlap measure on histograms, full width at half maximum of the histogram for the outside region of the segmented nodule on the background-corrected image, degree of irregularity, full width at half maximum for the inside region of the segmented nodule on the original image, and the contrast of the segmented nodule on the background-corrected image (23). LDA was trained with the same 150 cases, with locations of computer-detected true-positive (ie, lung cancer) and false-positive candidates (ie, benign nodules and non-nodules). Computer output indicated the likelihood of malignancy as a percentage, which was converted to a five-color scale, ie, red (80.0%–99.9%), orange (60.0%–79.9%), yellow (40.0%–59.9%), light green (20.0%–39.9%), and green (0.1%–19.9%).

To provide the likelihood of malignancy for any suspicious lesions that the radiologist might indicate interactively on a chest image in the observer study, we selected in advance all potential locations likely to be identified by radiologists. These locations were determined with locations of initially identified nodule candidates obtained by the computerized scheme for lung nodule detection. Mean number of initially identified nodule candidates was 67 for 150 cases. In the observer study, when the radiologist marked a point where the computer did not indicate, we identified the nearest location of the preestimated likelihood of malignancy for the location marked by the radiologist. The performance level of the computerized scheme for distinguishing malignant nodules from benign nodules and computer-detected false-positive candidates, which was determined by use of the same database for training and testing the computerized scheme, indicated that sensitivity and specificity were 87.7% and 66.7% (a threshold of 50% was used for the likelihood of being a lung cancer) when locations of all candidates were provided from the computerized scheme for nodule detection at the sensitivity level of 80.6%, with 1.2 false-positive results per image for the 108 lung nodules, respectively.

Observer Performance Study

We used ROC analysis to evaluate the performance of radiologists in the detection of lung cancers without and with computer output by use of a sequential test method (28). In the sequential test method, radiologists first were asked whether the cancer was present or absent, and they then marked their confidence level regarding the likeli-

hood of the presence of a cancer by using a continuous rating scale displayed on the monitor (29,30). After the radiologist marked the initial level of confidence, a number of computer outputs were shown on the chest image, and they again were asked to mark their confidence level if they wished to change the initial result. Nine radiologists, including five chest radiologists and four general radiologists, participated in this observer study.

In our observer study interface, locations of computer-detected nodule candidates were marked by arrowheads, with a five-color scale corresponding to different levels of likelihood of malignancy. In addition, radiologists were able to obtain the likelihood of malignancy for any suspected nodule candidates that could be located interactively by the radiologist, but were not detected by the computer. We used one monochrome (1600 × 1200 pixel; 800 cd/m^2; Totoku Electric Co Ltd, Tokyo, Japan) and one color liquid crystal display (LCD) monitor (1600 × 1200 pixel; 220 cd/m^2; Totoku Electric Co Ltd) for this observer performance study. We used the color LCD monitor for displaying computer output and other information for the observer study, and the high-contrast (600:1) monochrome LCD monitor, for interpreting chest images. Radiologists were allowed to change the window level and width on the LCD monitor. Reading time was not limited. Radiologists were asked about the presence of lung cancer on an image, and they then marked their confidence level regarding the likelihood of the presence of lung cancer by using a continuous rating scale (26) and also identified the location of the most suspicious lesion. It is important to note that radiologists were forced to select the most suspicious location for Localization ROC (LROC) analysis (36), even if they believed there was no lung cancer on the image. As soon as the radiologist marked the initial level of confidence and location, a large circle with one of five colors corresponding to the computer estimation of the likelihood of malignancy was displayed at the location the radiologist indicated. In addition, computer outputs of nodule candidates detected by the computer and color-coded likelihood of malignancy for each were displayed on the monitor with arrowheads. The radiologist then was asked again to mark his or her confidence level and location if he or she wished to change from the initial decision.

Before training and testing, radiologists were instructed as follows: (1) the purpose of this study is to evaluate the usefulness of CAD for lung cancer diagnosis on chest radiographs; (2) the role of CAD output is to serve as a "second opinion"; (3) 50% of 48 chest images include solitary lung cancer, and another 50% of chest images are non-nodule cases or include benign nodules; (4) the digitized chest image is shown on the monochrome LCD monitor on the right side; (5) CAD output is shown on the color LCD monitor with color markers indicating the likelihood of the lesion being lung cancer as follows: red, high probability of being lung cancer; yellow, moderate probability; and green, low probability; (6) computer performance for marking of lung nodules is about 80% sensitivity, with 1.2 false-positive results per image; (7) accuracy of computer output for the distinction of lung cancer and other benign nodules or false-positive results is about 85% sensitivity and 70% specificity when a threshold of 50% is used as the cutoff value for benign versus malignant; (8) click on a bar on the screen by using a mouse to indicate your confidence level regarding the presence or absence of a lung cancer, first without and then with computer output; (9) after each decision (without and with CAD), click one location by using the right button of a mouse; and (10) try to use the rating scale consistently and uniformly.

For a training session, we used another four cases (two lung cancer, one benign, and one non-nodule case) selected from the 150 cases. In the training session only, the actual diagnosis (ie, location of lung cancer) was indicated on the monitor after the radiologists' final decision based on the computer output.

ROC analysis was used for comparison of radiologists' performances in lung cancer diagnosis on chest radiographs without and with computer output. A binormal ROC curve was fitted to each radiologist's confidence-rating data from the two reading conditions with quasi-maximum likelihood estimation (37). A computer program (PROPROC, [Metz CE, The University of Chicago, IL] [38]) was used for obtaining binormal ROC and LROC curves for each radiologist. The statistical significance of the difference between ROC curves obtained without and with computer output was tested by use of the computer program, LABMRMC (Metz CE, The University of Chicago, IL)(39), which uses analyses of variance in pseudo-values of the area under the best fit binormal ROC curve (AUC) calculated from all rating scores of all radiologists (40). In addition, we applied JAFROC (Chakraborty DP, University of Pittsburgh, PA) software (41,42) to test the statistical significance of difference between two LROC data sets. JAFROC software was developed by Chakraborty and Berbaum (42) for estimating the statistically significant difference between two averaged free-response ROC (FROC) curves by use of the

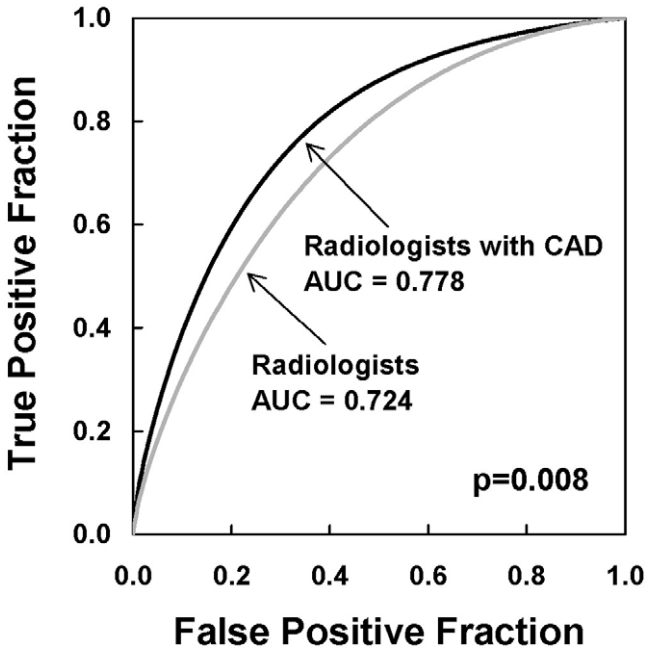

Figure 1. Average ROC curves obtained from nine radiologists for the detection of lung cancers without and with computer output.

jackknife method, which is the same as that used in the LABMRMC program (40). To apply the JAFROC program for testing LROC data sets obtained from our study, we assumed that LROC data sets would become comparable to free-response ROC results obtained under the specific condition in which the observer was informed that only one lesion was included in each case.

RESULTS

Figure 1 indicates ROC curves for the detection of lung cancers obtained from all radiologists without and with computer output. The average AUC of radiologists for the detection of lung cancers improved significantly ($P = .008$) from without (0.724) to with CAD (0.778). Table 2 lists AUC values without and with computer output for each radiologist. When the performance of all radiologists for the detection of lung cancers was divided into two subgroups, that of four general radiologists and that of five chest radiologists, the difference between average AUCs without and with CAD indicated statistical significance for only the general radiologists' group ($P = .010$) because the average gain in AUCs for the chest radiologists' group was relatively small (gain in AUC = 0.044) compared with the gain for the general radiologists (gain in AUC = 0.066).

Table 2
AUC Values for Nine Radiologists (four general and five chest radiologists) Without and With Computer Output

	AUC		
	Radiologist Only	With Computer	*P*
General radiologists			
A	0.729	0.782	
B	0.683	0.761	
C	0.685	0.743	
D	0.744	0.818	
Mean	0.710	0.776	.010*
Chest radiologists			
E	0.753	0.812	
F	0.837	0.877	
G	0.758	0.815	
H	0.735	0.762	
I	0.589	0.633	
Mean	0.734	0.778	.066
All radiologists mean	0.724	0.778	.008*

*Difference is statistically significant with 95% confidence.

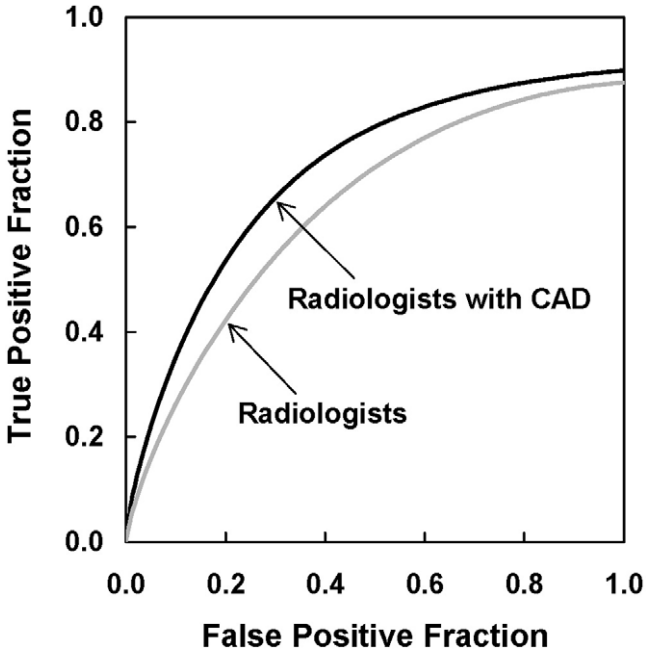

Figure 2. Average LROC curves obtained from nine radiologists for the detection of lung cancers without and with computer output.

In addition to average ROC curves, we obtained average LROC curves, as shown in Figure 2, to examine how radiologists detected lung cancers correctly in terms of their locations. There was a statistically significant difference ($P = .007$) between the two LROC data sets (0.690

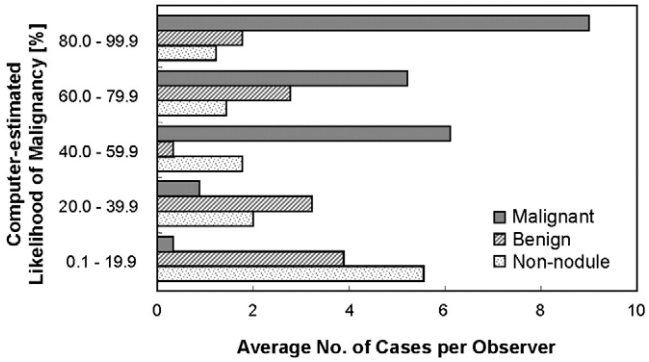

Figure 3. Mean number of correctly marked or not marked malignant and benign/non-nodule cases per observer at each level of likelihood of malignancy. Likelihoods were estimated by using the computerized scheme at the location of the most suspicious lesions marked by nine radiologists.

vs 0.742) for all radiologists. From LROC analysis, it should be noted that the mean number of lung cancers not detected correctly by all radiologists with CAD output was 2.44 (10.2%) of 24 lung cancers. However, of these false-negative cases, the mean number of lung cancers that was marked correctly by the computer was 2.11, but these were not considered to be lung cancer by the majority of radiologists; these cases therefore were considered to be "nonactionable" and extremely difficult lung-nodule cases (29) that were included in this observer study.

Figure 3 shows the relationship between mean number of "the most suspicious lesions" correctly identified by each radiologist for 24 malignant, 12 benign, and 12 non-nodule cases and the level of likelihood of malignancy provided to the radiologists in the observer study. We assumed the cancer was identified correctly if cancer location was marked correctly by the radiologist, regardless of how the radiologists rated the probability of lung cancer. For benign and normal cases, we assumed that all responses for benign and normal cases were identified correctly because we required all radiologists to mark one location even if they recognized that there was no lung cancer. Please note that results in Figure 3 include only one data point per nodule/non-nodule case even if more than one computer output was displayed on the monitor. If we assume that the computer output for the most suspicious lesion for a radiologist was considered as correct, ie, "true positive," when output for malignant cases was at the level of 4 (60.0%–79.9%) or 5 (80.0%–99.9%), average sensitivity for the 24 malignant cases was 61.6%. In the same way, if we assume "true negative" for benign cases when output was at the level of 1 (0.1%–19.9%) or 2 (20.0%–39.9%), average specificity was 62.0%. Because we assumed that computer output in the range of computer-estimated likelihood of malignancy between 40.0% and 60.0% would not be useful for radiologist decision making, we did not include this range of cutoff values in calculating sensitivity and specificity.

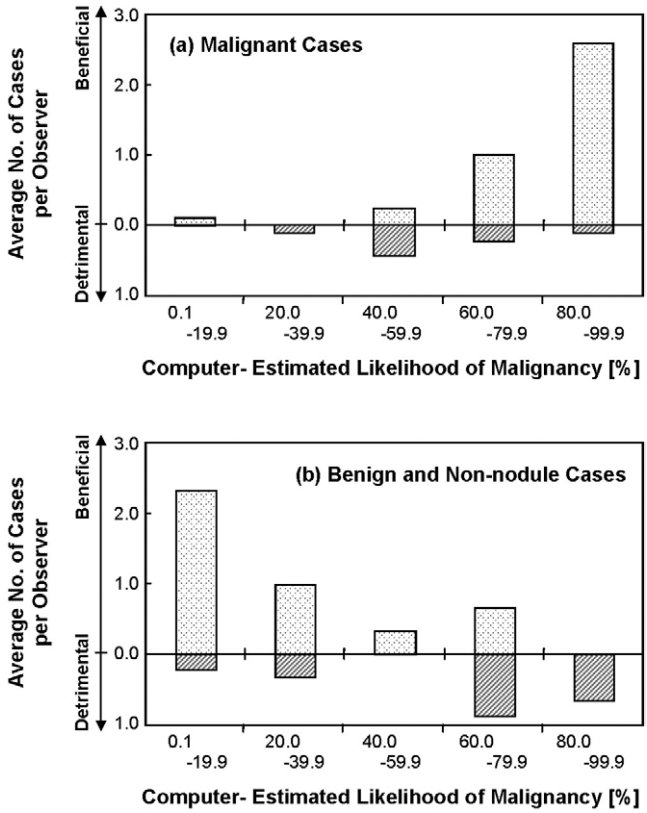

Figure 4. Mean number of correctly marked cases in which ratings were changed beneficially or detrimentally (threshold value in the difference, 0.1) for **(a)** 24 malignant and **(b)** 12 benign and 12 non-nodule cases between two readings by nine radiologists without and with computer output. The computer-estimated likelihood of malignancy in this figure was obtained at the location of the most suspicious lesions marked by nine radiologists.

Figure 4 shows the mean number of correctly identified cases per observer caused by beneficial and detrimental changes between two readings without and with computer output for (1) the 24 malignant cases and (2) the 12 benign and 12 non-nodule cases. The result in Figure 4 was obtained for only one nodule/non-nodule per case, ie, the same most suspicious lesions, as those in Figure 3. We used 0.10 of the difference between two confidence level ratings without and with computer output as the threshold for determining beneficial and detrimental changes for each radiologist. It is apparent from Figure 4 that high and low computer-estimated likelihoods of ma-

Table 3
Number of Beneficial and Detrimental Changes by Use of Computer Output for Nine Radiologists in 24 Malignant and 12 Benign and 12 Non-nodule Cases

	24 Malignant Cases			12 Benign and 12 Non-nodule Cases		All 48 Cases		
Radiologist	Beneficial	Detrimental	Correctly Marked	Beneficial	Detrimental	Beneficial	Detrimental	Correctly Marked
A	5	0	21	4	4	9	4	45
B	5	0	19	4	2	9	2	43
C	8	1	21	5	4	13	5	45
D	3	0	23	1	0	4	0	47
E	3	5	23	14	2	17	7	47
F	4	2	24	5	2	9	4	48
G	5	0	20	4	2	9	2	44
H	1	0	22	1	2	2	2	46
I	1	0	21	1	1	2	1	45
Mean	3.9	0.9	21.6	4.3	2.1	8.2	3.0	45.6
P		.005*		.132			.011*	

*Difference is statistically significant with 95% confidence.

lignancy were beneficial to radiologists for malignant cases and benign and non-nodule cases, respectively.

When we applied Student paired *t*-test to the number of cases considered beneficial and detrimental by all radiologists, there were statistically significant differences between beneficial and detrimental changes for all correctly identified cases ($P = .011$; mean number of correctly marked cases was 45.6) and the mean of 21.6 correctly marked malignant cases ($P = .005$), but not for 24 benign/non-nodule cases ($P = .132$), as listed in Table 3. These results indicate that the accuracy of radiologist decision making for lung cancer diagnosis improved significantly in malignant cases with use of computer output.

DISCUSSION

Several previous studies showed that radiologist performance improved significantly by use of computer aids for the detection (27–29,43) and also classification of lung nodules (30) on chest radiographs. However, these observer studies were carried out independently for the detection or classification of lung nodules, such that the effect of radiologist performance was studied only for each isolated task. For example, in an observer performance study for the detection of lung nodules (29), radiologists focused on detecting extremely subtle or very subtle lung nodules regardless of malignant/benign classification. Conversely, in an observer study for classification of lung nodules on chest radiographs (30), lung nodule locations were provided in advance so the radiologist would be able to focus his or her attention on the classification task without and with computer output, even if some lung nodules might be considered "nonactionable."

It would be reasonable to assume that the gain in radiologist performance provided by CAD for the classification task would be very small if lung nodules were not detected initially by radiologists. In this study, we therefore evaluate radiologist performance in lung cancer diagnosis in which their task should be considered the combination of two tasks, detection and classification of lung nodules, by use of the advanced CAD scheme.

Because computation time of the likelihood of malignancy for each candidate was too long and impractical for real-time computation, we estimated the likelihood of malignancy in all potential locations for all cases in advance. Therefore, locations provided to the computerized scheme for lung nodule classification were not always identical to those provided by radiologists in the observer study. However, we believe the number of locations for classification was sufficient for covering almost all lung fields; thus, the difference between the real location marked by the radiologist and the location of the preestimated candidate would be negligibly small. Moreover, if computer performance will be increased in terms of central processing unit speed, we can obtain the likelihood of malignancy in real time at the point identified by radiologists.

Results obtained from this observer study indicate that the overall performance of radiologists for the detection and classification of lung nodules on chest images improved significantly. The beneficial changes achieved by use of computer output were greater than the detrimental changes for both malignant and benign/non-nodule cases. Although we may assume that the main role of chest radiographs for lung cancer diagnosis is to detect suspicious lesions, rather than classify them, radiologist performance in lung cancer diagnosis could be improved by a system that indicates which lesions were suspicious. In this study, the number of lung cancers detected correctly by radiologists increased only slightly by use of computer output, probably because a few extremely subtle lung cancers were included. We believe these extremely subtle lung cancers were considered to be "nonactionable" because the majority of radiologists did not recognize these candidates as "lung cancer" in the observer study, even if the computer provided a correct location for some extremely subtle cancers. Therefore, it appears in this study that the improvement in radiologist performance in lung cancer diagnosis was provided mainly by the computer-estimated likelihood of malignancy. However, we believe the CAD scheme for detection of lung nodules also would contribute to improvement in lung cancer diagnosis in clinical practice, where a larger numbers of various "actionable" nodules would be encountered.

We designed this observer study to evaluate radiologist performance in a somewhat realistic situation in the sense that detection and classification tasks were combined; however, there are several limitations. First, only one lung nodule was allowed to be included in each case; this is considered one of the fundamental limitations in ROC analysis. In most clinical situations, the number of lesions is unknown at the time of interpretation, and more than one lesion commonly could be included in one image. Therefore, we may use free-response ROC analysis (42,44), rather than conventional ROC analysis, when we evaluate radiologist performance in the future.

The performance level of the CAD scheme used in this study was relatively high; this is a second limitation because performance level was evaluated by use of the same image database for training and testing CAD schemes. In practice, the performance level of a CAD system for detection of lung nodules on chest radiographs is likely to be somewhat less, such as 65.0% sensitivity with 5.3 false-positive results per image (45). Therefore, for our results to be applied to clinical situations, further improvement in the computerized scheme would be required.

Another limitation in this study is the small number of cases, 48, used in the observer study. This is because the time required for a radiologist has to be limited to 1 hour or so in one session. A large number of cases may be required if we want to include various types of lung nodules to investigate further the effects of radiologist performance in detection and classification tasks in the diagnosis of lung cancers on chest radiographs.

REFERENCES

1. Heelan RT, Flehinger BJ, Melamed MR, et al. Non-small-cell lung cancer: results of the New York screening program. Radiology 1984; 151:289–293.
2. Sone S, Takashima S, Li F, et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. Lancet 1998; 351:1242–1245.
3. Flehinger BJ, Kimmel M, Melamed MR. The effect of surgical treatment on survival from early lung cancer. Implications for screening. Chest 1992; 101:1013–1018.
4. Sobue T, Suzuki T, Matsuda M, Kuroishi T, Ikeda S, Naruke T. Survival for clinical stage I lung cancer not surgically treated. Comparison between screen-detected and symptom-detected cases. The Japanese Lung Cancer Screening Research Group. Cancer 1992; 69:685–692.
5. Miettinen OS. Screening for lung cancer. Radiol Clin North Am 2000; 38:479–486.
6. Muhm JR, Miller WE, Fontana RS, Sanderson DR, Uhlenhopp MA. Lung cancer detected during a screening program using four-month chest radiographs. Radiology 1983; 148:609–615.
7. 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.
8. Austin JH, Romney BM, Goldsmith LS. Missed bronchogenic carcinoma: radiographic findings in 27 patients with a potentially resectable lesion evident in retrospect. Radiology 1992; 182:115–122.
9. Mulshine J, Sullivan DC. Lung cancer screening. N Engl J Med 2005; 352: 2714–2720.
10. Gohagan JK, Marcus PM, Fagerstrom RM, et al. Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer. Lung Cancer 2005; 47:9–15.
11. Giger ML, Doi K, MacMahon H. 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.
12. Lampeter WA, Wandtke JC. Computerized search of chest radiographs for nodules. Invest Radiol 1986; 21:384–390.
13. Xu XW, Doi K, Kobayashi T, MacMahon H, Giger ML. Development of an improved CAD scheme for automated detection of lung nodules in digital chest images. Med Phys 1997; 24:1395–1403.
14. Lo SC, Freedman MT, Lin JS, Mun SK. Automatic lung nodule detection using profile matching and back-propagation neural network techniques. J Digit Imaging 1993; 6:48–54.
15. Carreira MJ, Cabello D, Penedo MG, Mosquera A. Computer-aided diagnoses: automatic detection of lung nodules. Med Phys 1998; 25: 1998–2006.
16. Giger ML, Bae KT, MacMahon H. Computerized detection of pulmonary nodules in computed tomography images. Invest Radiol 1994; 29: 459–465.
17. Lee Y, Hara T, Fujita H, Itoh S, Ishigaki T. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. IEEE Trans Med Imaging 2001; 20:595–604.
18. Armato SG III, Giger ML, Moran CJ, Blackburn JT, Doi K, MacMahon H. Computerized detection of pulmonary nodules on CT scans. Radiographics 1999; 19:1303–1311.

19. Armato SG III, Giger ML, MacMahon H. Automated detection of lung nodules in CT scans: preliminary results. Med Phys 2001; 28:1552–1561.
20. 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.
21. Arimura H, Katsuragawa S, Suzuki K, et al. Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening. Acad Radiol 2004; 11:617–629.
22. Nakamura K, Yoshida H, Engelmann R, et al. Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks. Radiology 2000; 214:823–830.
23. Aoyama M, Li Q, Katsuragawa S, MacMahon H, Doi K. Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images. Med Phys 2002; 29:701–708.
24. McNitt-Gray MF, Hart EM, Wyckoff N, Sayre JW, Goldin JG, Aberle DR. A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results. Med Phys 1999; 26:880–888.
25. Aoyama M, Li Q, Katsuragawa S, Li F, Sone S, Doi K. Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images. Med Phys 2003; 30:387–394.
26. Armato SG III, Altman MB, Wilkie J, et al. Automated lung nodule classification following automated nodule detection on CT: a serial approach. Med Phys 2003; 30:1188–1197.
27. Matsumoto T, Yoshimura H, Giger ML, et al. Potential usefulness of computerized nodule detection in screening programs for lung cancer. Invest Radiol 1992; 27:471–475.
28. Kobayashi T, Xu XW, MacMahon H, Metz CE, Doi K. Effect of a computer-aided diagnosis scheme on radiologists' performance in detection of lung nodules on radiographs. Radiology 1996; 199:843–848.
29. Shiraishi J, Abe H, Engelmann R, Doi K. Effect of high sensitivity in a computerized scheme for detecting extremely subtle solitary pulmonary nodules in chest radiographs: observer performance study. Acad Radiol 2003; 10:1302–1311.
30. Shiraishi J, Abe H, Engelmann R, Aoyama M, MacMahon H, Doi K. Computer-aided diagnosis to distinguish benign from malignant solitary pulmonary nodules on radiographs: ROC analysis of radiologists' performance—initial experience. Radiology 2003; 227:469–474.
31. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. Invest Radiol 1989; 24:234–245.
32. Shiraishi J, Katsuragawa S, Ikezoe J, et al. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. AJR Am J Roentgenol 2000; 174:71–74.
33. Xu XW, Doi K. Image feature analysis for computer-aided diagnosis: detection of right and left hemidiaphragm edges and delineation of lung field in chest radiographs. Med Phys 1996; 23:1613–1624.
34. Xu XW, Doi K. Image feature analysis for computer-aided diagnosis: accurate determination of ribcage boundary in chest radiographs. Med Phys 1995; 22:617–626.
35. Wu YC, Doi K, Giger ML, Metz CE, Zhang W. Reduction of false positives in computerized detection of lung nodules in chest radiographs using artificial neural networks, discriminant analysis, and a rule-based scheme. J Digit Imaging 1994; 7:196–207.
36. Starr SJ, Metz CE, Lusted LB, Goodenough DJ. Visual detection and localization of radiographic images. Radiology 1975; 116:533–538.
37. Metz CE, Herman BA, Shen JH. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. Stat Med 1998; 17:1033–1053.
38. Pan X, Metz CE. The "proper" binormal model: parametric receiver operating characteristic curve estimation with degenerate data. Acad Radiol 1997; 4:380–389.
39. Metz CE. LABMRMC software. Accessed at: http://www-radiology.uchicago.edu/krl/roc_soft.htm.
40. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol 1992; 27:723–731.
41. Zheng B, Chakraborty DP, Rockette HE, Maitz GS, Gur D. A comparison of two data analyses from two observer performance studies using Jackknife ROC and JAFROC. Med Phys 2005; 32:1031–1034.
42. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. Med Phys 2004; 31:2313–2330.
43. Kakeda S, Moriya J, Sato H, et al. Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system. AJR Am J Roentgenol 2004; 182:505–510.
44. Chakraborty DP, Winter LH. Free-response methodology: alternate analysis and a new observer-performance experiment. Radiology 1990; 174:873–881.
45. Freedman MT, Lo SCB, Osicka T, et al. Computer aided detection of lung cancer on chest radiographs: Effect of machine CAD false positive locations on radiologists' behavior. Proc SPIE Medical Imaging: Image Processing 2002; 4684:1311–1319.