Effect of CAD on Radiologists' Detection of Lung Nodules on Thoracic CT Scans: Analysis of an Observer Performance Study by Nodule Size

Berkman Sahiner, Ph.D.[1], **Heang-Ping Chan, Ph.D.**[1], **Lubomir M. Hadjiiski, Ph.D.**[1], **Philip N. Cascade, M.D.**[1], **Ella A. Kazerooni, M.D.**[1], **Aamer R. Chughtai, M.D.**[1], **Chad Poopat, M.D.**[2], **Thomas Song, M.D.**[2], **Luba Frank, M.D.**[1], **Jadranka Stojanovska, M.D.**[1], and **Anil Attili, M.D.**[1]

[1] Department of Radiology, The University of Michigan, 1500 East Medical Center Drive, Ann Arbor, MI 48109

[2] Department of Radiology, Henry Ford Hospital, 2799 W. Grand Boulevard, Detroit, MI 48202-2608

Abstract

Rationale and Objectives—To retrospectively investigate the effect of a computer aided detection (CAD) system on radiologists' performance for detecting small pulmonary nodules in CT examinations, with a panel of expert radiologists serving as the reference standard.

Materials and Methods—Institutional review board approval was obtained. Our data set contained 52 CT examinations collected by the Lung Image Database Consortium, and 33 from our institution. All CTs were read by multiple expert thoracic radiologists to identify the reference standard for detection. Six other thoracic radiologists read the CT examinations first without, and then with CAD. Performance was evaluated using free-response receiver operating characteristics (FROC) and the jackknife FROC analysis methods (JAFROC) for nodules above different diameter thresholds.

Results—241 nodules, ranging in size from 3.0 to 18.6 mm (mean 5.3 mm) were identified as the reference standard. At diameter thresholds of 3, 4, 5, and 6 mm, the CAD system had a sensitivity of 54%, 64%, 68%, and 76%, respectively, with an average of 5.6 false-positives (FPs) per scan. Without CAD, the average figures-of-merit (FOMs) for the six radiologists, obtained from JAFROC analysis, were 0.661, 0.729, 0.793 and 0.838 for the same nodule diameter thresholds, respectively. With CAD, the corresponding average FOMs improved to 0.705, 0.763, 0.810 and 0.862, respectively. The improvement achieved statistical significance for nodules at the 3 and 4 mm thresholds (p=0.002 and 0.020, respectively), and did not achieve significance at 5 and 6 mm (p=0.18 and 0.13, respectively). At a nodule diameter threshold of 3 mm, the radiologists' average sensitivity and FP rate were 0.56 and 0.67, respectively, without CAD, and 0.67 and 0.78 with CAD.

Conclusion—CAD improves thoracic radiologists' performance for detecting pulmonary nodules under 5 mm on CT examinations, which are often overlooked by visual inspection alone.

Corresponding Author: Berkman Sahiner, Ph.D., MIB C480A, 1500 East Medical Center Drive, Ann Arbor, MI 48109-5842, Phone: 734-647-7429, Fax: 734-615-5513, berki@umich.edu.

Keywords

Lung Nodule; CT; Computer-aided detection

INTRODUCTION

Although there is controversy over whether screening with computed tomography (CT) may reduce lung cancer mortality (1,2), there is little doubt that thin-slice multi-detector row CT (MDCT) allows the detection of more lung nodules, often of smaller size, than both thicker section CT examinations and chest radiographic images (3–6). MDCT examinations of the thorax are now commonly reconstructed at 1–3 mm slice thickness and often with overlapping intervals (7–9), resulting in a large number of images to be interpreted. Fatigue caused by the increased workload of reviewing more images per examination, combined with pressures to be time-efficient, may result in false-negatives (i.e., missed lung nodules). Computer-aided detection (CAD) may help reduce false-negatives at a lower cost than double reading by a second radiologist.

With increased interest in the detection and evaluation of lung nodules on CT examinations, a number of research groups have been investigating the development of CAD systems for pulmonary nodules (10–19), and the effect of CAD on radiologists' performance (20–23). Several CAD systems have been approved to be used as a "second reader" for applications in mammography and lung imaging (24,25). In a second reader paradigm, the radiologist first interprets the image without CAD, and then re-interprets it after the CAD prompts are shown (26). If the radiologist does not discard his/her previous findings in the re-interpretation, the nodule detection sensitivity and the number of false-positives (FPs) when reading with CAD will not be lower than that when reading without CAD. It is important to measure the potential increase in FPs with CAD that may adversely impact the clinical management, as well as the potential increase in sensitivity. Since many cases contain multiple nodules, it is also important to evaluate the effect of CAD on a per-nodule basis, as opposed to a per-scan only basis. Finally, with the goal of detecting cancer at an earlier stage, it is important to analyze the effect of CAD based on nodule size. To address these issues, we designed a free-response receiver operating characteristic (FROC) experiment to compare the detection of lung nodules on CT examinations with and without CAD, and to analyze the dependence of the performance results on nodule size.

MATERIALS AND METHODS

Data Set

Our data set consisted of two groups. Group 1 was obtained from the Lung Image Database Consortium (LIDC), the cases in which were collected by the participating institutions with IRB approval, and Group 2 was collected with IRB approval from patient files at our institution. Our study was Health Insurance Portability and Accountability Act (HIPAA) compliant. The criteria for the inclusion of a CT examination in the LIDC data set are described in the literature (27). Briefly, these CT examinations were collected from both diagnostic and screening studies at the five institutions participating in the LIDC. The cases were intended to capture the complexities of CT scans encountered during routine clinical practice, and to represent a variety of technical parameters from a variety of scanner models. Cases containing abnormalities other than lung nodules were not excluded, unless the other abnormality substantially interfered with visual interpretation. The upper limit on the number of nodules per scan was set at six (27). As of the beginning of our study, the LIDC data set contained annotated examinations from 73 patients. Although it is expected that the LIDC cases would include both screening and diagnostic studies (27) and contain both malignant and benign nodules, the number of cases

containing malignant nodules and the proportion of screening and diagnostic examinations were not available from the LIDC archive. The tube current for the Group 1 scans ranged from 40 to 486 mA.

Group 2 of the data set consisted of 91 CT examinations from patient files at our institution. Eighty of the Group 2 cases were selected from lung CT examinations of patients between 55 and 74 years of age, with a current or former smoking history of at least 30 pack years, performed in a 12 month period at our institution between Oct. 2002 and Sept. 2003. Sixty-six of these 80 scans were randomly selected from files of patients whose clinical reports for the scan indicated at least one non-calcified lung nodule, and 14 were randomly selected from those that did not indicate any non-calcified lung nodules. The reported prevalence of lung cancer in similar high-risk populations detected in initial CT screening varied from about 1% to 3% in recent studies (9,28–30). To include proven cancer cases in our data set, we enriched this group with 11 CT examinations that were randomly selected from patients with biopsy-proven malignant lung nodules. The tube current for these 11 scans from Group 2 ranged from 160 to 360 mA, and that for the remaining 80 scans ranged from 40 to 160 mA. The data set in this study therefore contained a larger proportion of malignant cases than what would be found in a screening population. The pool for the data set for this study thus consisted of 164 scans from 164 patients. The combination of the two groups of cases was needed to obtain a large enough sample size that satisfied the selection criteria for the observer study.

As explained below, each of these 164 CT examinations was read by multiple expert thoracic radiologists to establish a reference standard. The reference standard was established in two phases. All 164 examinations were included in the first phase, described in more detail below, which took place before the observer study. Based on the data collected in the first phase, a subset of cases was selected to be included in the observer study. For inclusion in the observer study, we required that a case 1) did not contain more than six nodules detected in the first phase with size ≥ 3 mm; 2) did not contain cases with more than two calcified nodules; and 3) contained at most ten micro-nodules, i.e., nodules less than 3 mm in size. These three criteria were used so that the duration of the observer study can be kept reasonably short, while the maximum amount of useful data can be collected. The limits for ≤ 6 nodules and ≤ 2 calcified nodules were set up so that cases with many nodules or calcified nodules did not constitute a major portion of the nodules present in the data set. The criterion for few micro-nodules aimed at reducing the reading time for the radiologists, since the purpose of the study was to investigate only the detection of nodules ≥ 3mm in diameter. Although no upper limit was set regarding nodule size for inclusion, nodules ≥30 mm in size were excluded from analysis, as described below in the Statistical Analysis Section. The attenuation characteristics of the nodules (e.g., solid vs. part solid and non-solid) were not part of the exclusion criteria. The final data set for the observer study selected with these criteria contained 85 CT examinations, with 49 CT examinations from Group 1 and 36 CT examinations from Group 2. Six CT examinations from Group 2 contained biopsy-proven malignant nodules. Pathology information about nodules from Group 1 was not provided by LIDC.

Table 1 lists the manufacturer and model of the CT scanners that acquired our final data set of 85 CT examinations, and the reconstruction filters used. In-plane resolution ranged from 0.49 to 0.82 mm (average = 0.67, median = 0.68 mm), and slice interval ranged from 1.25 to 3.0 mm (average = 2.15, median = 2.5 mm). Tube current ranged between 40 to 486 mA (median = 80 mA), and the anode potential ranged between 120 to 140 kVp (median = 120 kVp).

The reference standard

All CT examinations in our data set were read by multiple expert radiologists in two phases. The first phase took place before, and the second phase took place after the observer study. In the first phase, Group 1 cases were read by four experienced thoracic CT radiologists, and

Group 2 cases were read by two experienced thoracic CT radiologists. In the second phase, all cases that were included in the observer study were read in consensus by a panel of two experienced thoracic CT radiologists.

First phase—A detailed description of the truth assessment for Group 1 (LIDC) cases can be found in the literature (27,31); a brief summary of the relevant aspects are provided below. For this group, four experienced thoracic CT radiologists first read the cases independently. Readers marked the location of nodules ≥ 3 mm in diameter, and drew the boundary of the nodule on each slice on which it appeared. Nodules < 3 mm in diameter were marked only if they were not clearly benign. Locations of non-nodule objects that might be confused with nodules, such as scars, atelectasis, postsurgical changes, etc. were also marked if they were ≥ 3 mm in diameter. This first reading was termed the "blinded" reading pass. After blinded reading, each radiologist read all the cases again, this time in an unblinded reading session, in which the markings of all radiologists were available. In unblinded reading, the radiologists were free to modify the nodule locations they marked during blinded reading, add or delete objects, modify the boundaries of an annotation. They also subjectively characterized several properties of each nodule ≥ 3 mm.

All examinations in Group 2 were read independently by two experienced thoracic radiologists using a graphical user interface (GUI) developed in our laboratory in collaboration with radiologists for reading CT images and providing descriptors of clinical relevance. One of the two radiologists also served as one of the four expert readers in the LIDC. The expert radiologists had 18 and 30 years of experience in interpreting thoracic CT scans. They identified the location of any lung nodule that they detected, irrespective of nodule size, and enclosed the lesion with a rectangular 2D bounding box on the slice where it was best visualized. They also indicated the first and the last slice where the lesion could be seen. These two expert radiologists did not participate as readers in the observer study.

Second phase—In the second phase, two expert radiologists, each with over 15 years experience as thoracic radiologists, read all 85 cases included in the observer study as a consensus panel. These two radiologists were the same as those that read Group 2 cases in the first phase. In this phase, the radiologists did not search for nodules, but instead focused on regions of interest (ROIs) displayed on the CT volumes. These ROIs consisted of any suspicious region, regardless of size, identified by any radiologist either in the without-CAD or with-CAD modes in the observer study, or those identified by the expert radiologists in the first phase. The panel radiologists were blinded as to how or by whom a particular ROI was identified. After examining the ROI and the surrounding tissue in adjacent slices, and after a discussion on the characteristics of the finding if necessary, the two radiologists reached a consensus as to whether each ROI contained a nodule or not. If a finding was categorized as a nodule, the length of its longest axis and the perpendicular axis on the slice containing the largest axial cross-section of the nodule was measured using an electronic ruler. The long axis length of the nodule was used as the nodule diameter. The diameters of findings that were categorized as FPs were not measured. For nodules with a diameter of ≥ 3 mm, the radiologists provided ratings for the subtlety of the nodule relative to those encountered in clinical practice, and likelihood of malignancy, both on a scale of 1 to 5.

Computerized Nodule Detection System

The computer system was trained on a data set consisting of 94 chest CT examinations containing lung nodules that were collected from patient files at our institution (32). This training set is independent of the test set used in the observer study. The in-plane resolution of the training cases ranged from 0.546 mm to 0.969 mm, and the section thickness ranged from 1.0 mm to 5.0 mm. A block diagram of the computer system, consisting of prescreening,

segmentation, feature extraction, rule-based FP reduction, and LDA-based FP reduction stages, is shown in Figure 1. The nodule detection algorithm is summarized below.

In the prescreening stage, nodule candidates were detected using adaptive 3D clustering within the lung regions. A 3D active contour (AC) algorithm, initialized by k-means clustering within a 3D volume of interest (VOI) containing the object, was used for segmentation of the object (33). Three groups of features were extracted for each VOI. The first group of features was extracted by considering the statistics of the direction and magnitude of the gradient vector on spherical shells centered at each voxel of the detected nodule candidate (34). Nineteen features were extracted by using statistics (e.g., standard deviation, average, coefficient of variation) of the gradient field direction and magnitude. The second group was related to eigenvalues of the Hessian matrix, which are derived from the second-order derivatives within the VOI (32, 34). In order to reduce the noise in the second-order derivatives that are used to compute the Hessian matrix, the image is first smoothed using a Gaussian filter bank with different spatial widths (or filter scales) of the Gaussian filter,. Hessian eigenvalues from different scales are combined using an artificial neural network (ANN) trained to distinguish between pixels belonging to a nodule and those belonging to normal structures. For a given test VOI, the ANN outputs at different voxels were combined to define a Hessian feature for the VOI (35). The third group of features was related to the size and shape of the object, and included the volume and the surface area of the segmented 3D object, as well as a combination of slice-based features for each object, such as the area, perimeter, circularity, compactness, major and minor axes and their ratio, and eccentricity of a fitted ellipse in each cross section of the 3D object. Six ellipsoid features were also extracted by fitting a 3D ellipsoid to the segmented object, and using the lengths and the length ratios of the three principal axes. A complete list of our size and shape features can be found in the literature (11,34).

Our FP reduction method consisted of three steps. First, each object was automatically classified as either juxta-pleural (JP) or internal. Second, rule-based classification was applied to both the JP and internal subsets, based on the Hessian feature described above. Rule-based classification was designed to limit the number of JP and internal objects to 25 and 20 per examination, respectively. Finally, two classifiers, one trained for JP objects and one trained for internal objects were used for FP reduction. Linear discriminant analysis (LDA) classifiers with stepwise feature selection were designed using the training set for this final FP reduction stage. After training, all parameters were fixed and the CAD system was applied to the set of 85 test CT scans. The CAD marks were stored on the display workstation before the observer study.

Observer Performance Study

Six study radiologists participated as observers, different from the two expert radiologists who read the cases in phase 1 for the establishment of the reference standard. Four were fellowship-trained in cardiothoracic radiology, and two were cardiothoracic radiology fellows in the last few months of their fellowship year. The study radiologists read each case first without and then with CAD, referred to as without- and with-CAD modes, respectively. A GUI, shown in Figure 2, was designed in our laboratory for the study. It allows the user to scroll through the CT slices efficiently in cine-mode, mark the ROIs, and record the ratings. By default, the images were first displayed using a standard lung window; the radiologists had the option to adjust the window and level settings to improve lesion visibility at any time. There was no limit on the reading time. The study radiologists were not informed of the proportion of examinations with and without lung nodules, nor the maximum number of reference nodules for a case in the data set. They were instructed to mark any objects that they considered to be a nodule, regardless of its size or likelihood of malignancy. To keep the length of the observer study to a reasonable duration, the study radiologists did not measure the diameter of their findings. A training

session with five CT examinations was conducted for each observer to allow them to get familiar with the GUI and the reading process of the experiment. The training cases were different from the 85 CT examinations in the test data set.

In the without-CAD mode, the radiologists searched the image volume for nodules in a manner similar to their clinical reading. Following written instructions for the reading session, they provided a mark-rating pair for each suspected nodule by enclosing the central slice of the suspected lesion with a rectangle, and providing a likelihood of being a nodule (LN) rating on a scale of 1 to 100. They also provided likelihood of malignancy and nodule subtlety ratings, although these two ratings were not used for data analysis in the current study. Immediately following the without-CAD reading, the GUI displayed the CAD prompts indicating the computer-detected regions. A computer-detected region was shown as a rectangle enclosing the object on its central slice. The radiologist examined the slices around the CAD mark, and either rejected the CAD mark as an FP, or accepted it and provided an LN rating. For the radiologist's reference, suspected nodules detected by himself/herself in the without-CAD mode were also displayed as rectangles in a different color. If the computer-detected region coincided with the study radiologist's marking in the without-CAD mode, the radiologist's LN rating for the mark was retrieved and displayed.

If a radiologist-detected finding is not detected by the computer, the miss is not considered to be a reason for the radiologist to ignore the finding or to lower the level of suspicion for it in the with-CAD mode, especially for CAD systems with a moderate sensitivity. In addition, if radiologists are allowed to focus on areas of the image that are not prompted by the CAD system in the with-CAD reading mode, it may be difficult to distinguish whether the improvement, if any, is due to the use of CAD, or due to having a second opportunity to read the image. For these reasons, in the with-CAD mode of our study, the radiologists were allowed to act upon CAD prompts only. For findings that were detected by the study radiologist and missed by the computer, the LN rating in the with-CAD mode was defined to be the same as that in the without-CAD mode. As a result, with CAD, the number of marks per image either increased or remained the same.

Statistical Analysis

The mark-rating pairs were analyzed using FROC methodology (36), which uses the location and the ratings for each reported abnormality to relate the sensitivity to the number of FPs per examination. Jackknife FROC analysis method (JAFROC) was used to statistically compare the radiologists' performance in the without- and with-CAD modes. We used the JAFROC-1 software, which was shown, through simulation studies, to have a higher statistical power than alternative analysis methods for human observers (including human-observers with CAD) (37). JAFROC-1 software uses the FP marks on both normal and abnormal images, in contrast to an earlier version of the JAFROC software that uses the FP marks on only normal images. The software provides a figure-of-merit (FOM) for each reader that is related to the probability that a rating provided for a reference nodule exceeds that for an FP, and an overall p-value for the difference between the two reading modes based on the mark-rating pairs of all readers. The FROC curves shown in the Results Section plotted the sensitivity with respect to the number of FPs per examination (FP rate) averaged over all scans.

We analyzed the differences between the without- and with-CAD modes at four nodule diameter thresholds, namely D=3 mm, 4 mm, 5 mm, and 6 mm. In addition to JAFROC analysis, we also computed the sensitivity and the average FP rate for each radiologist at these four different nodule diameter thresholds. For the analysis of sensitivity and FP rate, we used an LN threshold of 0 as an indication of no detection, i.e., any finding detected by a radiologist, regardless of the level of suspicion, was included in the calculation of the sensitivity and FP

rate. The statistical significance of the difference in sensitivities or FP rates between without- and with-CAD was estimated with the use of the Student's paired t-test for the 6 radiologists.

The first step in scoring the radiologists' marks is to determine if they correspond to nodules of interest. As described above, our data set contained small nodules (< 3 mm) and masses. The detection of a small nodule by a radiologist should not be penalized as an FP, but cannot be considered as a true-positive (TP) either because these nodules may be too small to be of clinical significance. Similarly, we were not interested in the detection of masses (lesions with diameter ≥ 3 cm) that would be very unlikely to be missed in clinical interpretation. We therefore labeled marks that coincided with a small nodule or a mass as marks of no consequence (ONC). The definition of what constitutes a TP, depended on the nodule diameter threshold. For a nodule diameter threshold of D, the mark was labeled as a TP if the diameter of the overlapping reference nodule was larger than or equal to D, and the mark was labeled as an ONC if the diameter of the overlapping reference nodule was less than D. All other marks were labeled as FPs.

RESULTS

Figure 3 shows the distribution of the nodule diameter for our data set of 85 CT examinations. The sizes of nodules of interest varied between 3.0 mm and 18.6 mm, with a mean of 5.3 mm and a median of 4.4 mm. There were a total of 241, 142, 85, and 54 reference nodules with diameter ≥ 3, 4, 5, and 6 mm, respectively. Figure 4 shows the distribution of the number of nodules per examination. Twenty-one CT examinations did not contain any nodules greater than or equal to 3 mm in diameter, while there was at least one nodule in the remaining 64 scans.

The computer detection system had a sensitivity of 54% (130/241) 64% (91/142), 68% (58/85), and 76% (41/54), with an average of 5.6 FPs per examination at a diameter threshold of 3, 4, 5, and 6 mm, respectively (Table 2). The FROC curves of the six radiologists without- and with-CAD for a diameter threshold of 3 mm are shown in Figure 5. The FOM indicated by JAFROC analysis in the without- and with-CAD modes at this nodule diameter threshold are shown in Table 3 for each radiologist. The average FOM and the 95% confidence interval (in parentheses) were 0.661 (0.622–0.698) and 0.705 (0.664–0.743) in the without- and with-CAD modes, respectively, as estimated by the JAFROC-1 software.

Table 4 shows the corresponding average FOM values for different diameter threshold values. The average FOM for the with-CAD mode was larger than that for the without-CAD mode for all nodule diameter thresholds. The improvement with CAD was statistically significant at diameter thresholds of 3 and 4 mm (p=0.002 and 0.020, respectively), and did not achieve significance at diameter thresholds of 5 and 6 mm (p=0.18 and 0.13, respectively).

Table 5 shows the average sensitivity and the average number of FP marks corresponding to the rightmost point in each FROC curve, i.e., if one uses an LN cutoff threshold of 0 to decide the presence of a nodule. The average sensitivity and the range for the six radiologists are shown for each diameter threshold. The potential dependence of the number of FP marks on the nodule diameter threshold was not studied because the sizes of the objects judged to be FPs were not measured by the consensus panel. The use of CAD increased the average sensitivity at a nodule diameter threshold D = 3 from 0.56 to 0.67 (18.9%). The increases at D = 4, 5, and 6 mm were 13.1%, 6.9% and 7.7%, respectively. The increase in sensitivity was significant at all thresholds studied (p<0.03, two-tailed Student's paired t-test). Accompanying with the increase in sensitivity, the average number of FPs per scan also increased significantly (p=0.003) from 0.67 to 0.78 (16.8%).

Considering that the computer had an average of 5.6 FPs per examination, the small, albeit significant increase in the FP rate indicates that the radiologists were efficient in rejecting computer FP marks. However, some TP marks by the computer were also rejected. Averaged over the six study radiologists, there were 9.0 (range: 3–16) reference nodules in the data set that were missed by a radiologist in the without-CAD mode, cued by the computer in the with-CAD mode, but rejected by the radiologist as an FP. Some of the computer-detected nodules were rejected by multiple radiologists. Figure 6 shows a 6 mm reference nodule that was dismissed as an FP by two radiologists in the observer study. In total there were 27 reference nodules identified by the CAD system that were rejected by at least one radiologist. Of these, 15 were larger than 4 mm, and 3 were larger than 6 mm in diameter.

Examples in Figures 7 to 9 further illustrate the effect of the CAD system. A beneficial effect of a TP CAD mark is illustrated in Fig. 7. In the without-CAD mode, this 4 mm nodule was detected by only 2 of the 6 study radiologists. When they inspected the CAD prompt in the with-CAD mode, the remaining 4 radiologists also marked this finding as a nodule. A detrimental effect of an FP CAD mark is illustrated in Fig. 8. Apical scars were frequently misclassified as nodules by the CAD system, as shown in this example. The number of radiologists marking this scar as a nodule increased from 1 in the without-CAD mode to 3 in the with-CAD mode. Finally, a 5 mm ground glass opacity that was detected by 3 study radiologists and missed by the CAD system is shown in Fig. 9.

DISCUSSION

In the past few years, several studies have investigated the effect of CAD on radiologists' interpretation of thoracic CT examinations for the detection of pulmonary nodules. Awai et al. (20) used a data set of 50 CT examinations with 10 observers to study the effect of a CAD system. Five of the observers were radiologists who did not specialize in thoracic radiology and five were radiology residents. The results were analyzed using Alternative ROC (AFROC) (38) analysis, and a paired t-test for the area A_1 under the AFROC curve for each radiologist. Average A_1 values without and with CAD were 0.64 and 0.67, respectively, and the difference was statistically significant (p<0.01). Li et al. (21) used a data set of twenty-seven low-dose CT studies including 17 patients with missed peripheral lung cancers. Fourteen observers including three thoracic radiologists read the cases without and with CAD, either in cine-mode or in a multiformat display, indicated the location of the most suspicious position on each CT examination, and provided a confidence level regarding the presence of lung cancer. ROC analysis indicated a significant improvement in A_z with CAD. Localization ROC (LROC) analysis also showed improvement; however, no statistical test was performed due to the lack of a standard method for statistical comparison of LROC curves. Brown et al. (22) used a data set of 6 CT examinations containing lung nodules, 2 nodule-free examinations, and a population of 202 observers to evaluate the effect of CAD. JAFROC analysis of the results of the 13 observers who read all eight cases indicated an improvement of 22% in FOM with CAD, but the improvement did not achieve statistical significance likely due to the limited sample size. Both the sensitivity and the FP rates of the 13 observers increased significantly with CAD. Das et al. (23) compared the effectiveness of two commercial CAD systems in assisting three radiologists with 1, 3, and 6 years of interpreting CT examinations of the chest, on a data set of 116 small lung nodules with a mean diameter of 3.4 mm. All radiologists showed improvement in their sensitivities with CAD, however, the FP rates without and with CAD was not analyzed.

Our study included a relatively large number of CT examinations (N=85) and used free-response ROC methodology for data collection, in which the observers provided a mark-rating pair for each perceived abnormality. The results were analyzed using JAFROC methodology, which has been shown, through simulation studies, to have higher statistical power compared

to ROC analysis (39). The statistics presented in our Results section may therefore be more robust than those that may be obtained using only a comparison of sensitivities, ROC or AFROC, or analysis with a smaller sample size.

Our results show that radiologists' sensitivity increased with the nodule diameter threshold, both with or without CAD. The FOM obtained by JAFROC analysis also showed the same trend. This is not surprising since larger nodules are easier to detect by both the radiologist and the CAD system. Radiologists' average sensitivity in the without-CAD mode increased from 0.56 to 0.84 as D increased from 3 to 6 mm. The CAD system's sensitivity alone was lower than the radiologists' average sensitivity at all nodule size thresholds. The difference Δ(sens) between the sensitivities of the radiologists and the CAD system increased as the nodule size threshold increased, with Δ(sens)=0.02, 0.03, 0.09, and 0.08, at D=3, 4, 5, and 6 mm, respectively. As a result, the improvement in radiologists' sensitivity with CAD was higher for D=3 and 4 mm than D=5 and 6 mm (Table 5). The percentage increase in sensitivity with CAD was 18.9%, 13.1%, 6.9% and 7.7% at D=3, 4, 5 and 6 mm, respectively. Accompanying the increase in sensitivity, we observed a statistically significant increase in the number of false-positives (Table 5), which could result in unnecessary additional workup and/or follow-up if the results were extrapolated to clinical practice. The increase in FOM with CAD showed a downward trend as D increased (Table 4), similar to that of the sensitivity change with CAD. As a result of this trend, and also as a result of the reduced sample size as D increased, the effect of CAD on the detection accuracy (FOM) was statistically significant for D=3 and 4 mm, and did not achieve significance for D=5 and 6 mm (Table 4).

Table 3 indicates that, at a nodule diameter threshold of D=3 mm, the radiologists with the two lowest FOM in without-CAD mode (R2 and R3) showed the largest improvement with CAD (ΔFOM=0.049 and 0.085 for R2 and R3, respectively). The radiologists with the highest FOM in without-CAD mode (R6) showed the second-smallest improvement with CAD (ΔFOM=0.023). The difference between the highest and the lowest FOM among the six radiologists was 0.17 in the without-CAD mode and 0.15 in the with-CAD mode. Similar trends were observed for the sensitivity. For example, at the same nodule diameter threshold of D=3 mm, the difference between the highest and the lowest sensitivity was 0.31 in the without-CAD mode and 0.21 in the with-CAD mode (Table 5). These results suggest that CAD may be useful for reducing the difference in lung nodule detection performance between radiologists.

As discussed in the Results section, the study radiologists rejected as FPs some of the reference nodules that were detected by the CAD system. Figure 6 shows a reference nodule that was marked by three of the four LIDC expert radiologists in unblinded reading mode, and was detected by the CAD system. In our observer study, this reference nodule was marked by one radiologist in the without-CAD mode. In the with-CAD mode, three additional study radiologists accepted the CAD mark as a nodule while two study radiologists rejected the CAD mark as an FP. The rejection of the prompts by one expert and two study radiologists is clearly a result of inter-radiologist variation in interpretation, rather than a failure to detect the abnormality. It has been argued that an important problem with CAD utilization is that radiologists frequently ignore true lesions detected by the CAD system (40,41). When the CAD system generates a fairly large number of FPs, radiologists' confidence in the system may be reduced and they may be more reluctant to accept computer prompts (42). Eye-tracking studies (43) suggest that the majority of missed lung cancers on chest radiographs are due to interpretation, rather than detection errors. Improvement of the FP rate of the CAD system, and alternative methods to present computer results (44) may help reduce interpretation errors.

The collection of a large data set with ground truth for an observer study is time-consuming. To reduce the data collection effort, we included in our data set cases that were already collected by the LIDC from five different institutions and annotated by expert radiologists. As discussed

above and shown in Table 1, our data set included scans acquired by CT scanners from different manufacturers, and different image acquisition and reconstruction parameters. The use of a heterogeneous test data set provided the advantage of assuring that the CAD system performance or the observer study results are not limited to a particular image acquisition or reconstruction protocol. In order to reduce heterogeneity in the reference standard, cases both from the LIDC and our institution were read in consensus by two experienced radiologists in a second-phase review to establish the reference standard.

We did not stratify the analysis of the effect of CAD by manufacturer, model, or reconstruction kernel type because of the relatively small sub-group sizes. However, we observed that the CAD system performance depended strongly on the reconstruction kernel type. For scans reconstructed using a smooth or medium-smooth kernel (standard, B31s or B30f) the average number of FPs per scan was 6.42, and for those reconstructed using a sharp kernel (bone or D) the average number of FPs per scan was 3.18. The sensitivities for the former and latter groups were 0.54 and 0.52, respectively, indicating a slightly lower sensitivity with a sharp kernel. These results are consistent with those in an earlier study (45) that compared CAD system performance for a data set of thoracic CT scans reconstructed using four different reconstruction kernels, which found that a smoother kernel resulted in a significantly higher FP rate and slightly higher sensitivity. These observations may be attributed to the more homogeneous gray-level values and smoother shapes of the objects segmented from CT scans reconstructed with smoother kernels, which may cause some of the non-nodule objects to be misclassified as nodules. Higher signal-to-noise ratio and better computer detectability are expected for low-contrast nodules with smoother reconstruction kernels (46).

Our study had a number of limitations. First, the number of malignant nodules was too small for statistical analysis of the effect of CAD on malignant and benign subgroups, as pathology information about nodules in Group 1 was not available, and Group 2 included only six cases that contained biopsy-proven malignant nodules. The results of our study for lung nodule detection therefore cannot be directly extrapolated to reveal the impact of CAD for lung cancer detection. However, differentiation of benign and malignant lung nodules on CT is difficult, especially for small nodules. The current guidelines for the management of small pulmonary nodules detected on CT scans from the Fleischner Society recommend that any non calcified lung nodule found on CT in a high-risk patients should be followed-up with CT at a minimum, with more aggressive management as nodule size increases (47). The additional nodules detected by radiologists using CAD could potentially result in earlier lung cancer detection, regardless of the appearance of the nodule on the current examination.

A second limitation of our study, related to the first, is that the reference standard of the presence of a nodule was established by a consensus panel rather than biopsy-proven. Given that small pulmonary nodules are generally not assessed by biopsy or resection, the reference standard was defined by an expert panel in several previous studies involving CAD for lung nodule detection on CT (16,23,48). Results of studies that rely on an expert panel, including our study, clearly depend on how the expert panel review is conducted, and how the "truth" is labeled based on that review. It has been shown in the literature that depending on how the reference standard was established based on the panel's readings, different modalities could be shown to be statistically significantly better than others (49). An analysis of the dependence of our results on the various combinations of "truth" obtained in the first and second phases used in the construction of the reference standard would be an interesting topic for future study.

A third limitation of our study is that some of the examinations had a large number of reference nodules. The detection of additional nodules with CAD may not make a difference in the clinical management of these patients. In our study design, we attempted to limit the number of such cases by excluding CT examinations with a large number of reference nodules detected

in the first phase. At the end of the second phase, due to additional nodules that were detected during the second phase, six examinations were determined to have more than seven nodules with diameter ≥3 mm (Fig. 4). The main conclusions of the study, however, do not necessarily depend on the number of nodules per examination, and are unlikely to be affected by these examinations.

We used the reference radiologists' readings for the diameter of TPs to analyze the size dependence retrospectively because the study radiologists did not measure the diameter of their findings. For FPs, there were no size measurements by the reference radiologists. As a result, when we compared radiologists' performance with and without CAD with respect to nodule size, all FPs were included regardless of the nodule size threshold. In a scenario in which the radiologists might be asked to indicate only nodules larger than a certain threshold, some of these FPs would not have been marked because their diameter might be smaller than the threshold. Therefore, our FROC curves and FOMs may be pessimistically biased, i.e., might be lower than they would have been if the radiologists were asked to indicate only nodules larger than a certain threshold. Since the with-CAD mode included more FPs, it might contain even more below-threshold FPs than those in the without-CAD mode. The pessimistic bias, if any, will therefore be larger in the with-CAD mode, which means that the trend for the improved performance with CAD will not be affected. In future studies, if the resources of the study permit, it may be beneficial to ask the study radiologists to measure the size of their findings.

We conclude that the use of a CAD system with a reasonable performance level has the potential to improve radiologists' performance in detecting pulmonary nodules in MDCT examinations, particularly smaller nodules which are more often missed by even thoracic radiology fellowship trained readers.

Acknowledgments

References

1. Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS. Survival of patients with stage I lung cancer detected on CT screening. New England Journal of Medicine 2006;355:1763–1771. [PubMed: 17065637]

2. Bach PB, Jett JR, Pastorino U, Tockman MS, Swensen SJ, Begg CB. Computed tomography screening and lung cancer outcomes. JAMA - Journal of the American Medical Association 2007;297:953–961.

3. Buckley JA, Scott WW, Siegelman SS, et al. Pulmonary nodules: effect of increased data sampling on detection with spiral CT and confidence in diagnosis. Radiology 1995;196:395–400. [PubMed: 7617851]

4. Ko JP, Naidich DP. Lung nodule detection and characterization with multislice CT. Radiologic Clinics of North America 2003;41:575–597. [PubMed: 12797607]

5. Sone S, Yang ZG, Takashima S, et al. Characteristics of small lung cancers invisible on conventional chest radiography and detected by population based screening using spiral CT. British Journal of Radiology 2000;73:137–145. [PubMed: 10884725]

6. Fischbach F, Knollmann F, Griesshaber V, Freund T, Akkol E, Felix R. Detection of pulmonary nodules by multislice computed tomography: improved detection rate with reduced slice thickness. Eur Radiol 2003;13:2378–2383. [PubMed: 12743736]

7. Cagnon CH, Cody DD, McNitt-Gray MF, Seibert JA, Judy PF, Aberle DR. Description and implementation of a quality control program in an imaging-based clinical trial. Academic Radiology 2006;13:1431–1441. [PubMed: 17111584]

8. Xu DM, van Klaveren RJ, de Bock GH, et al. Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. European Journal of Radiology 2008;68:347–352. [PubMed: 17920800]

9. Veronesi G, Bellomi M, Mulshine JL, et al. Lung cancer screening with low-dose computed tomography: A non-invasive diagnostic protocol for baseline lung nodules. Lung Cancer 2008;61:340–349. [PubMed: 18308420]

10. Reeves AP, Kostis WJ. Computer-aided diagnosis for lung cancer. Radiologic Clinics of North America 2000;38:497–509. [PubMed: 10855257]

11. Gurcan MN, Sahiner B, Petrick N, et al. Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. Medical Physics 2002;29:2552–2558. [PubMed: 12462722]

12. Wormanns D, Fiebich M, Saidi M, Diederich S, Heindel W. Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system. European Radiology 2002;12:1052–1057. [PubMed: 11976846]

13. Brown MS, Goldin JG, Suh RD, McNitt-Gray MF, Sayre JW, Aberle DR. Lung micronodules: Automated method for detection at thin-section CT - Initial experience. Radiology 2003;226:256–262. [PubMed: 12511699]

14. Suzuki K, Armato SG, Li F, Sone S, Doi K. Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography. Med Phys 2003;30:1602–1617. [PubMed: 12906178]

15. Armato SG, Roy AS, MacMahon H, et al. Evaluation of automated lung nodule detection on low-dose computed tomography scans from a lung cancer screening program. Academic Radiology 2005;12:337–346. [PubMed: 15766694]

16. Bae KT, Kim J-S, Na Y-H, Kim KG, Kim J-H. Pulmonary Nodules: Automated Detection on CT Images with Morphologic Matching Algorithm--Preliminary Results. Radiology 2005;236:286–293. [PubMed: 15955862]

17. Lee JW, Goo JM, Lee HJ, Kim JH, Kim S, Kim YT. The potential contribution of a computer-aided detection system for lung nodule detection in multidetector row computed tomography. Investigative Radiology 2004;39:649–655. [PubMed: 15486524]

18. Yuan R, Vos PM, Cooperberg PL. Computer-aided detection in screening CT for pulmonary nodules. American Journal of Roentgenology 2006;186:1280–1287. [PubMed: 16632719]

19. Rubin GD, Lyo JK, Paik DS, et al. Pulmonary nodules on multi-detector row CT scans: Performance comparison of radiologists and computer-aided detection. Radiology 2005;234:274–283. [PubMed: 15537839]

20. Awai K, Murao K, Ozawa A, et al. Pulmonary nodules at chest CT: Effect of computer-aided diagnosis on radiologists' detection performance. Radiology 2004;230:347–352. [PubMed: 14752180]

21. Li F, Arimura H, Suzuki K, et al. Computer-aided Detection of Peripheral Lung Cancers Missed at CT: ROC Analyses without and with Localization. Radiology 2005;237:684–690. [PubMed: 16244277]

22. Brown MS, Goldin JG, Rogers S, et al. Computer-aided Lung Nodule Detection in CT Results of Large-Scale Observer Test. Academic Radiology 2005;12:681–686. [PubMed: 15935966]

23. Das M, Muhlenbruch G, Mahnken AH, et al. Small Pulmonary Nodules: Effect of Two Computer-aided Detection Systems on Radiologist Performance. Radiology 2006;241:564–571. [PubMed: 17057074]

24. Bazzocchi M, Mazzarella F, Del Frate C, Girometti R, Zuiani C. CAD systems for mammography: a real opportunity? A review of the literature. Radiologia Medica 2007;112:329–353. [PubMed: 17440698]

25. Chan HP, Hadjiiski LM, Zhou C, Sahiner B. Computer-Aided Diagnosis of Lung Cancer and Pulmonary Embolism in Computed Tomography—A Review. Academic Radiology 2008;15:535–555. [PubMed: 18423310]

26. Taylor SA, Charman SC, Lefere P, et al. CT colonography: Investigation of the optimum reader paradigm by using computer-aided detection software. Radiology 2008;246:463–471. [PubMed: 18094263]

27. Armato SG, McLennan G, McNitt-Gray MF, et al. Lung Image Database Consortium: Developing a Resource for the Medical Imaging Research Community. Radiology 2004;232:739–748. [PubMed: 15333795]

28. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early lung cancer action project: overall design and findings from baseline screening. The Lancet 1999;354:99–105.

29. Swensen SJ, Jett JR, Hartman TE, et al. Lung cancer screening with CT: Mayo Clinic experience. Radiology 2003;226:756–761. [PubMed: 12601181]

30. Diederich S, Wormanns D, Semik M, et al. Screening for early lung cancer with low-dose spiral CT: Prevalence in 817 asymptomatic smokers. Radiology 2002;222:773–781. [PubMed: 11867800]

31. McNitt-Gray MF, Armato SG, Meyer CR, et al. The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation. Academic Radiology 2007;14:1464–1474. [PubMed: 18035276]

32. Sahiner B, Hadjiiski LM, Chan HP, Zhou C, Wei J. Computerized lung nodule detection on screening CT scans: performance on juxta-pleural and internal nodules. Proc SPIE 2006;6144:5S1–5S6.

33. Way TW, Hadjiiski LM, Sahiner B, et al. Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours. Medical Physics 2006;33:2323–2337. [PubMed: 16898434]

34. Ge Z, Sahiner B, Chan HP, et al. Computer aided detection of lung nodules: false positive reduction using a 3D gradient field method and 3D ellipsoid fitting. Medical Physics 2005;32:2443–2454. [PubMed: 16193773]

35. Sahiner B, Hadjiiski LM, Chan HP, et al. The effect of nodule segmentation on the accuracy of computerized lung nodule detection on CT scans: Comparison on a data set annotated by multiple radiologists. Proc SPIE 2007;6514:65140L–65141. 65147.

36. Bunch PC, Hamilton JF, Sanderson GK, Simmons AH. A free response approach to the measurement and characterization of radiographic observer performance. J Appl Photo Engr 1978;4:166–171.

37. Chakraborty DP. Validation and statistical power comparison of methods for analyzing free-response observer performance studies. Academic Radiology 2008;15:1554–1566. [PubMed: 19000872]

38. Chakraborty DP, Winter LHL. Free-response methodology: Alternate analysis and a new observer-performance experiment. Radiology 1990;174:873–881. [PubMed: 2305073]

39. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. Medical Physics 2004;31:2313–2330. [PubMed: 15377098]

40. Nishikawa RM, Kallergi M, Orton CG. Computer-aided detection, in its present form, is not an effective aid for screening mammography. Medical Physics 2006;33:811–814. [PubMed: 16696454]

41. Karssemeijer N, Otten JD, Rijken H, Holland R. Computer aided detection of masses in mammograms as decision support. British Journal of Radiology 2006;79:S123–S126. [PubMed: 17209117]

42. Zheng B, Swensson RG, Golla S, et al. Detection and classification performance levels of mammographic masses under different computer-aided detection cueing environments. Academic Radiology 2004;11:398–406. [PubMed: 15109012]

43. Manning DJ, Ethell SC, TD. Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph. British Journal of Radiology 2004;77:231–235. [PubMed: 15020365]

44. Karssemeijer, N.; Hupse, A.; Samulski, M.; Kallenberg, M.; Boetes, C.; den Heeten, G. In: Krupinski, EA., editor. An interactive computer aided decision support system for detection of masses in mammograms; 9th International Workshop on Digital Mammography; Tucson, AZ: Springer-Verlag Berlin; 2008. p. 273-278.

45. Ochs R, Angel E, Boedeker K, et al. The influence of CT dose and reconstruction parameters on automated detection of small pulmonary nodules. Proc SPIE 2006;6144:5W1–5W8.

46. Boedeker KL, McNitt-Gray MF. Tradeoffs in noise, resolution, and dose with reconstruction filter in lung nodule detection in CT. Proc SPIE 2005;5745:695–703.

47. MacMahon H, Austin JHM, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. Radiology 2005;237:395–400. [PubMed: 16244247]

48. Marten K, Engelke C, Seyfarth T, Grillhosl A, Obenauer S, Rummeny EJ. Computer-aided detection of pulmonary nodules: influence of nodule characteristics on detection performance. Clinical Radiology 2005;60:196–206. [PubMed: 15664574]

49. Revesz G, Kundel HL, Bonitatibus M. The effect of verification on the assessment of imaging techniques. Investigative Radiology 1983;18:194–198. [PubMed: 6862810]
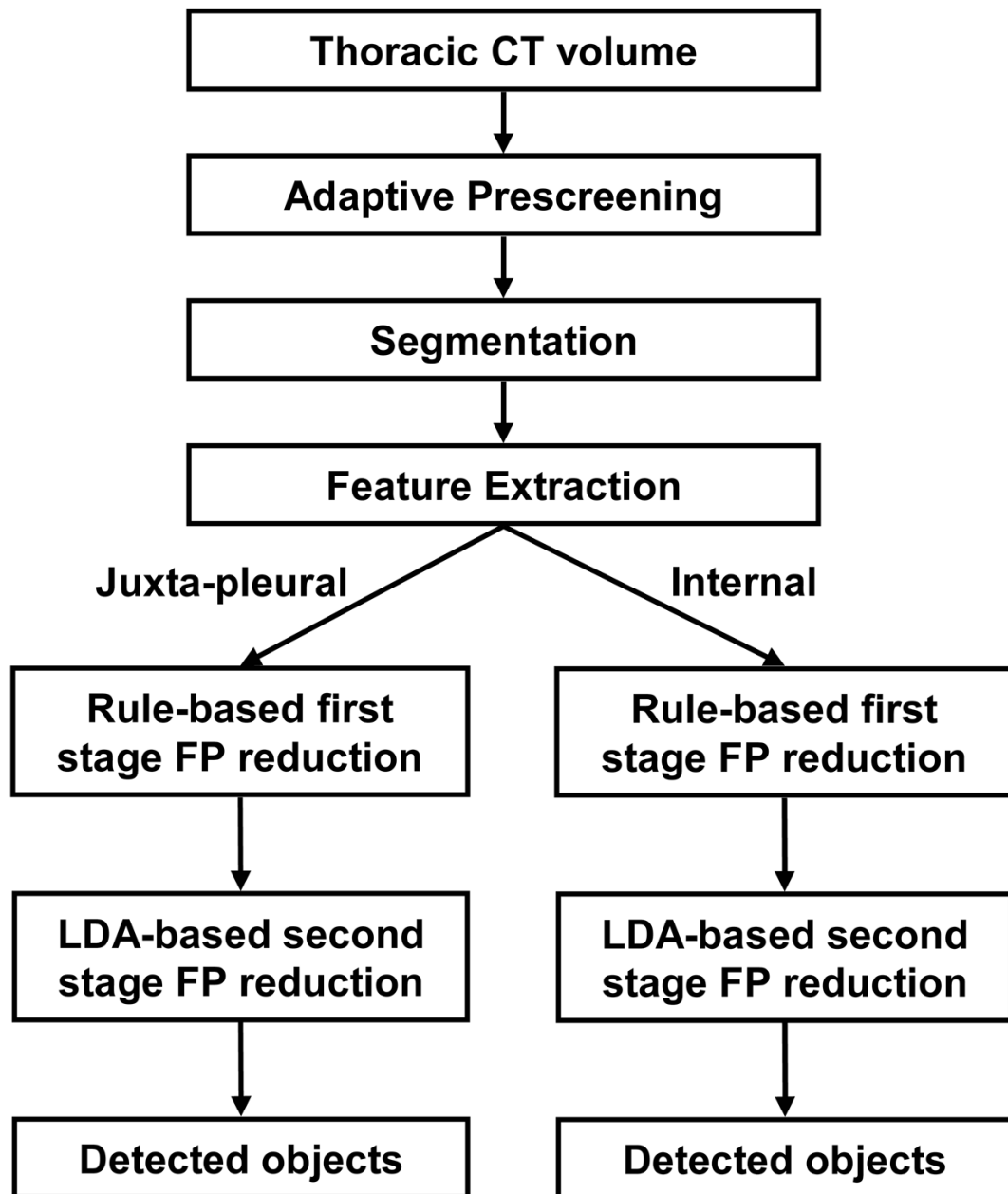
```
┌─────────────────────────┐
│    Thoracic CT volume   │
└─────────────────────────┘
              ↓
┌─────────────────────────┐
│   Adaptive Prescreening │
└─────────────────────────┘
              ↓
┌─────────────────────────┐
│       Segmentation      │
└─────────────────────────┘
              ↓
┌─────────────────────────┐
│    Feature Extraction   │
└─────────────────────────┘
```

Juxta-pleural   Internal

Rule-based first stage FP reduction	Rule-based first stage FP reduction
LDA-based second stage FP reduction	LDA-based second stage FP reduction
Detected objects	Detected objects

Figure 1.
The distribution of the nodule diameter for the 85 CT examinations in our data set.
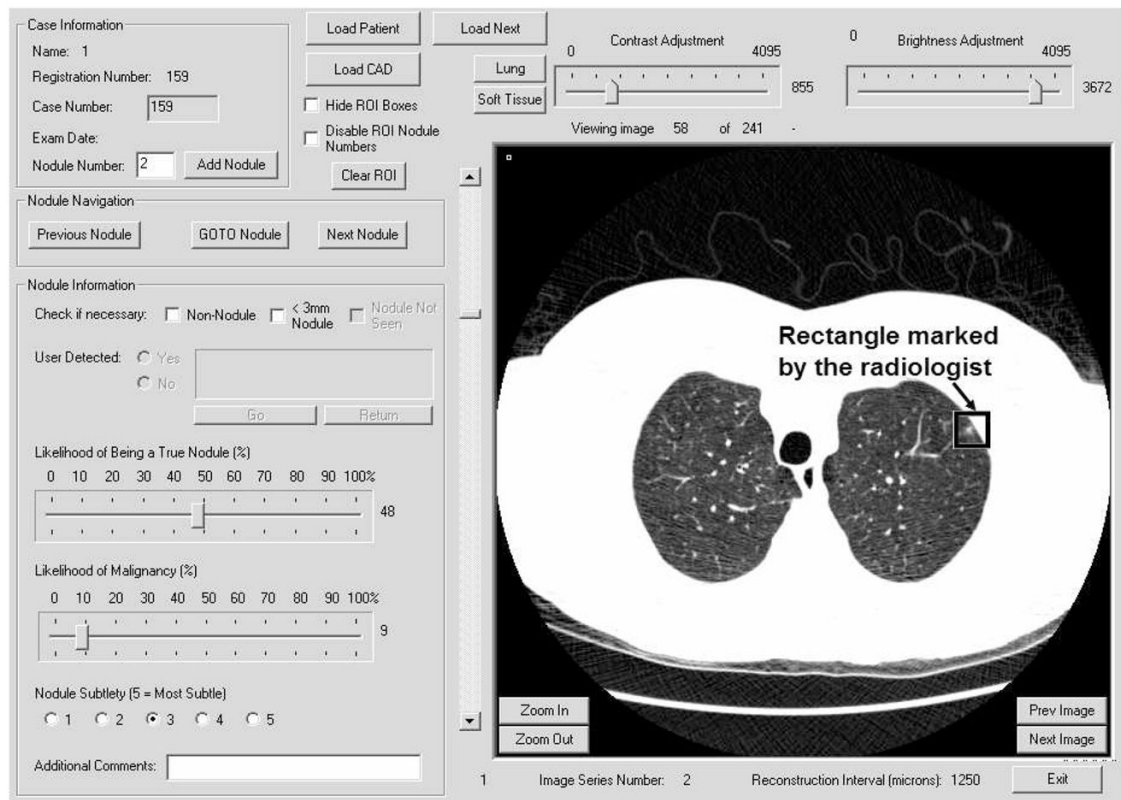
Figure 2.
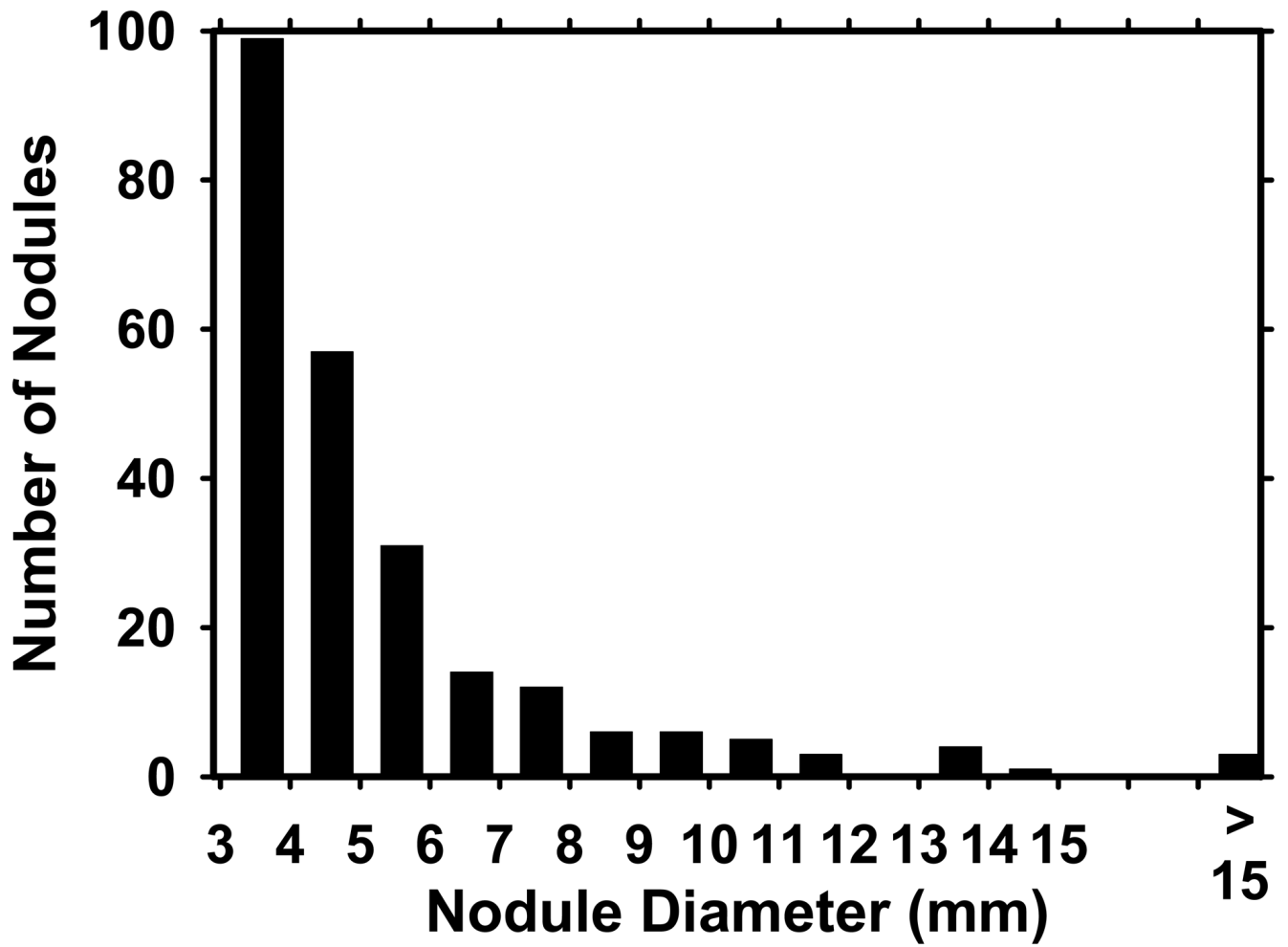The distribution of the number of lung nodules in a scan for the 85 CT examinations in our data set.

Figure 3.
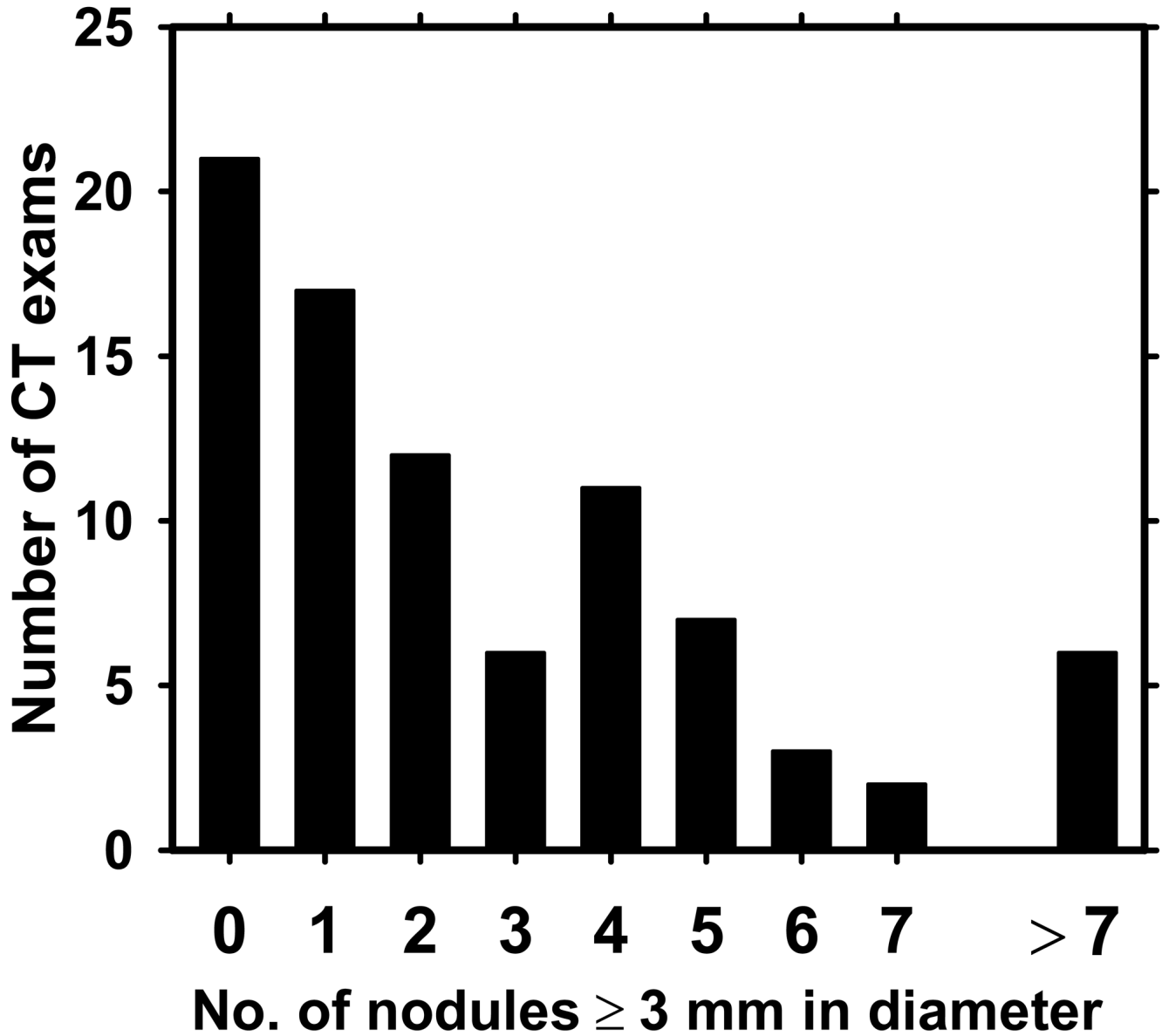The block diagram of our computerized nodule detection system.

Figure 4.
The GUI developed for the observer performance study. The figure shows an example of reading in the without-CAD mode, in which the radiologist provided a mark-rating pair for each suspected nodule by enclosing the central slice of the suspected lesion with a rectangle, and providing an LN rating on a scale of 1 to 100.
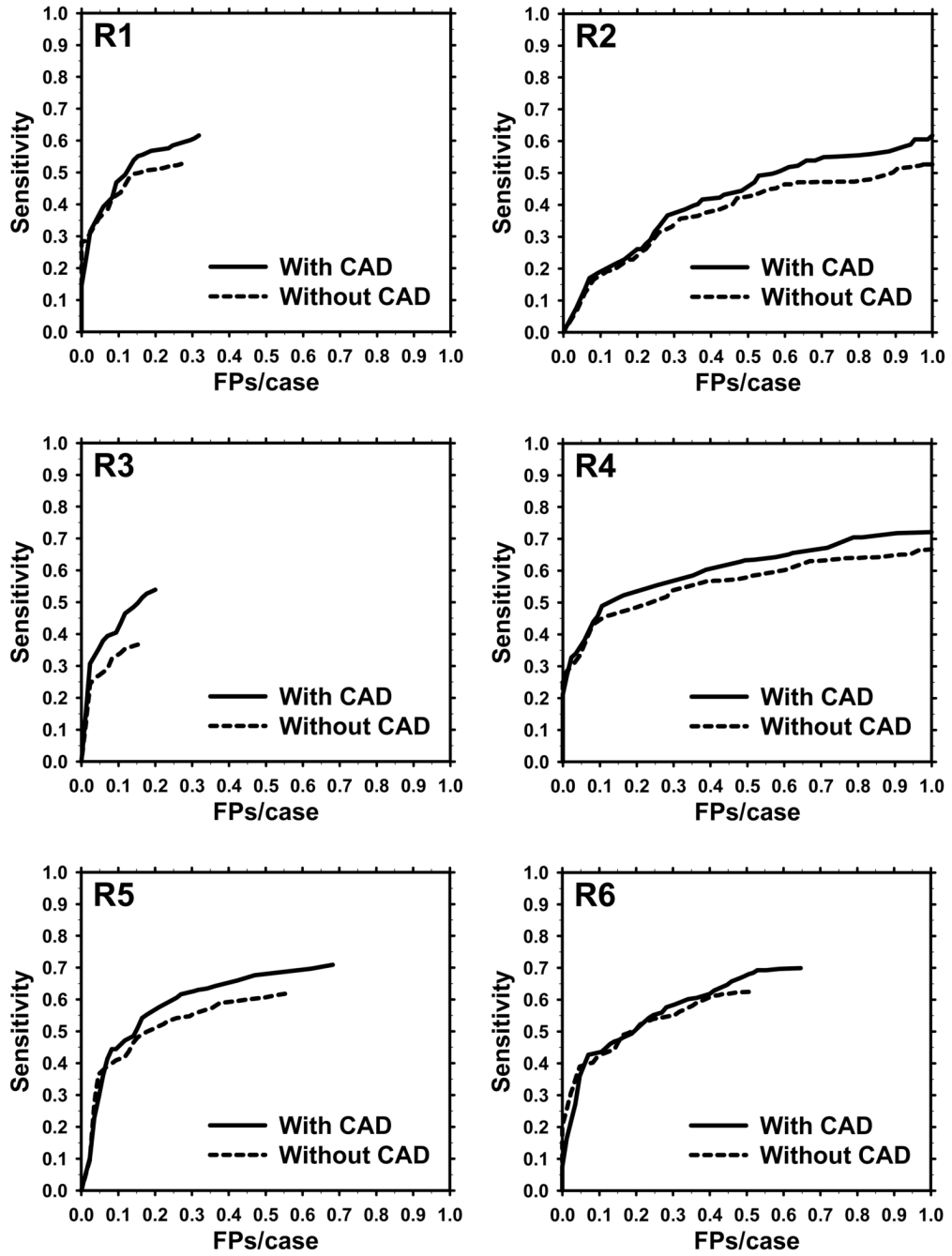
Figure 5.
FROC curves for the six study radiologists without and with CAD for nodules with diameter ≥ 4 mm (D=4).
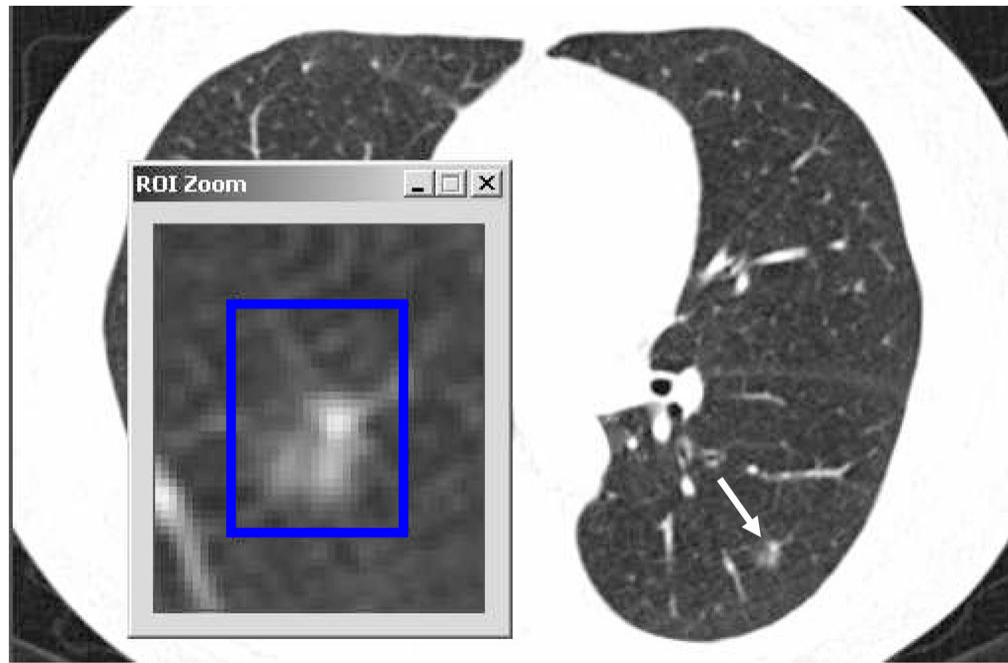
Figure 6.
A reference nodule that was detected by 1 and 4 study radiologists in the without- and with-CAD modes, respectively. Three of the four expert LIDC radiologists marked this reference nodule in the unblinded reading mode. This example demonstrates that some of the so-called misses in both without- and with-CAD modes may in reality be differences of interpretation between the study radiologists and the reference radiologists.
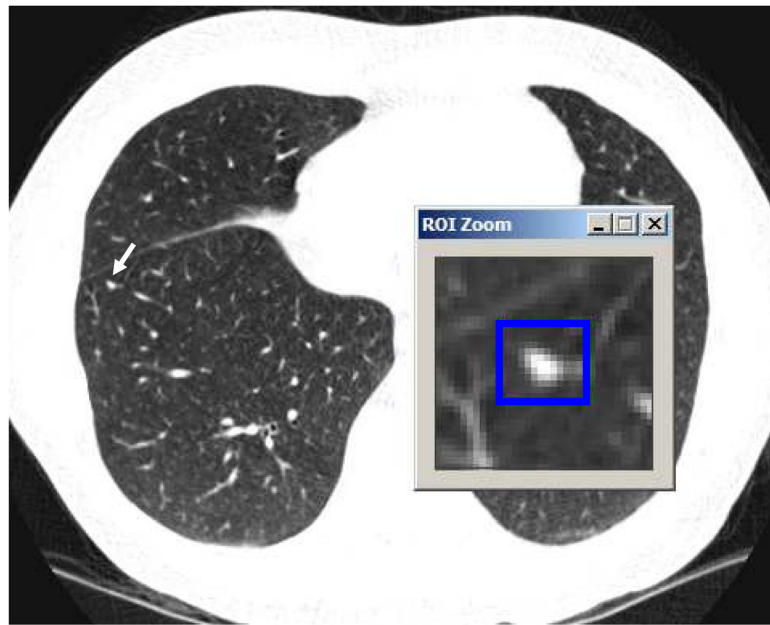
Figure 7.
A TP CAD mark that increased radiologists' sensitivity. This is a 4 mm reference nodule that was detected by 2 of the 6 study radiologists in the without-CAD mode. After viewing the computer prompt in the with-CAD mode, the other 4 study radiologists also marked this finding as a nodule.
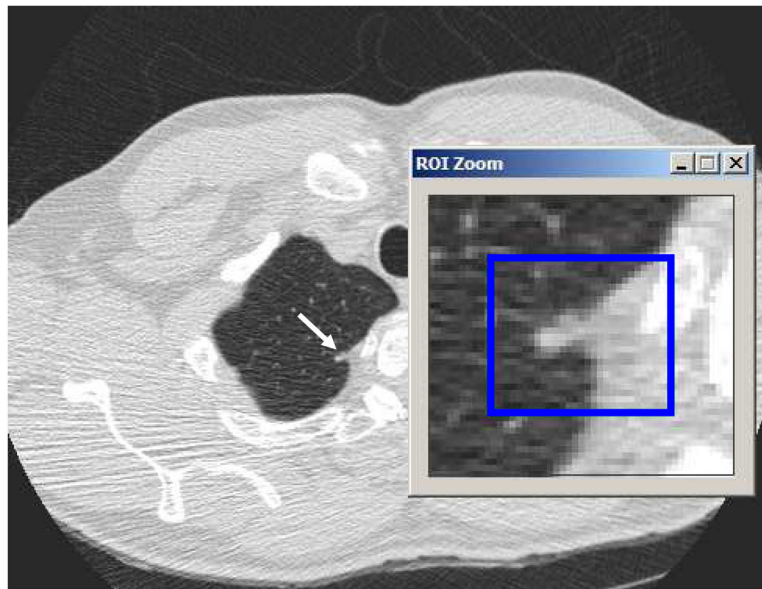
Figure 8.
An FP CAD mark that increased radiologists' FP rate. In the without-CAD mode, 1 out of 6 radiologists marked this apical scar as a nodule. After viewing the computer prompt in the with-CAD mode, two additional study radiologists marked it as a nodule.
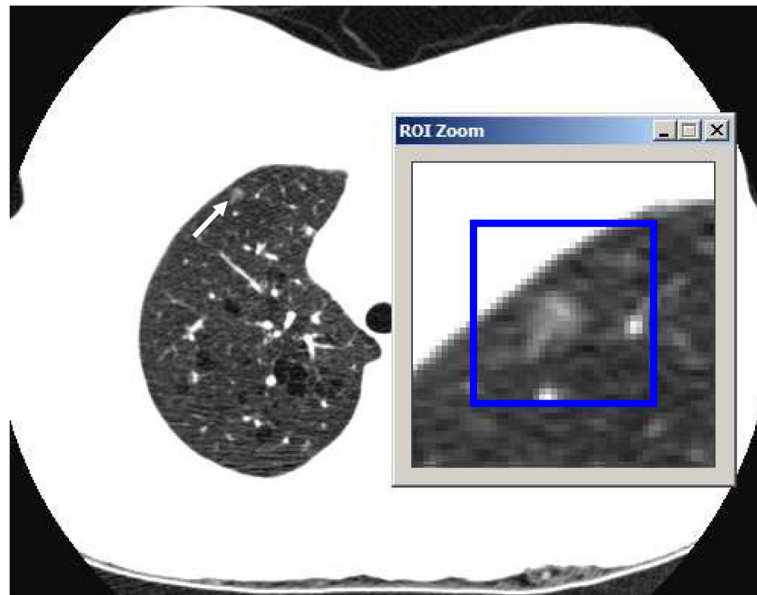
Figure 9.
An FN of the CAD system. This 5 mm ground-glass opacity was detected by 3 study
radiologists in the without-CAD mode, and missed by the CAD system.

Table 1

The Manufacturer and Model of the CT Scanners, and the Reconstruction Filter Type for the 85 Examinations in Our Data Set

Manufacturer/Model	Number of scans	Filter Type
GE LightSpeed QX/I	17	Bone
GE LightSpeed Ultra	1	Bone
GE LightSpeed Ultra	35	Standard
GE LightSpeed 16	5	Standard
Philips Brilliance 16	3	D[*]
Siemens Emotion 6	4	B31s[†]
Siemens Sensation 16	15	B30f[‡]
Siemens Sensation 64	5	B30f

[*] Filter D is described by the manufacturer as optimized for high-contrast objects

[†] Filter B31s is described by the manufacturer as medium-smooth filter

[‡] Filter B30f is described by the manufacturer as smooth filter

Table 2

Computer Detection Sensitivities for Nodule Diameter Thresholds of 3, 4, 5, and 6 mm, and the FP Rate.

	Average sensitivity at nodule diameter threshold			
FP rate (average # of FPs per scan)	3 mm	4 mm	5 mm	6 mm
5.6	54%	64%	68%	76%

Table 3

The Figure-of-Merit (FOM) by JAFROC Analysis in the Without- and With-CAD Modes at a Nodule Diameter Threshold of D=3 mm.

Radiologist	FOM without-CAD	FOM with-CAD	Increase with CAD
R1	0.697	0.736	0.039
R2	0.538	0.587	0.049
R3	0.628	0.713	0.085
R4	0.690	0.72	0.030
R5	0.706	0.741	0.035
R6	0.708	0.731	0.023
Average	0.661 (0.622–0.698)	0.705 (0.664–0.743)	0.044

Note.– Numbers in parentheses indicate the 95% confidence interval.

Table 4

Average FOM by JAFROC Analysis of the Six Study Radiologists for Nodule Diameter Thresholds of 3, 4, 5, and 6 mm.

Nodule Diameter Threshold (mm)	Average FOM without-CAD (95% confidence interval)	Average FOM with-CAD (95% confidence interval)	Increase with CAD	p-value
3	0.661 (0.662–0.698)	0.705 (0.664–0.743)	0.044	0.002
4	0.729 (0.685–0.770)	0.763 (0.717–0.805)	0.034	0.020
5	0.793 (0.743–0.836)	0.810 (0.759–0.853)	0.017	0.18
6	0.838 (0.782–0.884)	0.862 (0.808–0.905)	0.024	0.13

Table 5

Average FP Rate and Sensitivity of the Six Study Radiologists for Nodule Diameter Thresholds of 3, 4, 5, and 6 mm.

	FP rate (average # of FPs per scan)	Average sensitivity at nodule diameter threshold			
		3 mm	4 mm	5 mm	6 mm
without-CAD	0.667 (0.153–1.259)	0.559 (0.369–0.676)	0.670 (0.507–0.782)	0.773 (0.635–0.847)	0.836 (0.741–0.889)
with-CAD	0.778 (0.200–1.412)	0.665 (0.544–0.755)	0.758 (0.662–0.838)	0.825 (0.753–0.894)	0.901 (0.815–0.963)
% increase with CAD	16.8%	18.9%	13.1%	6.9%	7.7%
p-value	0.003	0.001	0.003	0.023	0.019

Note.– A likelihood of nodule (LN) threshold of 0 was used to determine which marking corresponded to a TP, an FP, or an ONC. Numbers in parentheses indicate the range. The statistical significances of the differences in sensitivities and FP rates between the without- and with-CAD modes were estimated with the Student's paired t-test for the 6 radiologists.