True Detection Versus “Accidental” Detection of Small Lung Cancer by a Computer-Aided Detection (CAD) Program on Chest Radiographs

Feng Li, Roger Engelmann, Kunio Doi, and Heber MacMahon

To evaluate the number of actual detections versus “accidental” detections by a computer-aided detection (CAD) system for small nodular lung cancers (≤ 30 mm) on chest radiographs, using two different criteria for measuring performance. A Food-and-Drug-Administration-approved CAD program (version 1.0; Riverain Medical) was applied to 34 chest radiographs with a “radiologist-missed” nodular cancer and 36 radiographs with a radiologist-mentioned nodule (a newer version 3.0 was also applied to the 36-case database). The marks applied by this CAD system consisted of 5-cm-diameter circles. A strict “nodule-in-center” criterion and a generous “nodule-in-circle” criterion were compared as methods for the calculation of CAD sensitivity. The increased sensitivities by the nodule-in-circle criterion were considered as nodules detected by chance. The number of false-positive (FP) marks was also analyzed. For the 34 radiologist-missed cancers, the nodule-in-circle criterion caused eight more cancers (24%) to be detected by chance, as compared to the nodule-in-center criterion, when using the version 1.0 results. For the 36 radiologist-mentioned nodules, the nodule-in-circle criterion caused seven more lesions (19%) to be detected by chance, as compared to the nodule-in-center criterion, when using the version 1.0 results, and three more lesions (8%) to be detected by chance when using the version 3.0 results. Version 1.0 yielded a mean of six FP marks per image, while version 3.0 yielded only three FP marks per image. The specific criteria used to define true- and false-positive CAD detections can substantially influence the apparent accuracy of a CAD system.

KEY WORDS: Lung, neoplasms, computer-aided detection, chest radiography

INTRODUCTION

The sensitivity of a computer-aided detection (CAD) system is an important criterion for determining its clinical utility. Many advanced noncommercial and commercial CAD systems have been developed for the detection of masses or microcalcifications on mammograms, and some

of them have been used for breast cancer screening programs.^{1–9} However, the performance of CAD schemes for detection of lung nodules on chest radiographs has been limited because of the typically high number of false-positive (FP) detections.^{10,11}

We have previously reported that the CAD sensitivity for the detection of radiologist-missed lung cancers was 35% (12/34) with 5.9 FP marks per radiograph with an Food and Drug Administration (FDA)-approved nodule detection CAD program (version 1.0, Riverain Medical, Miamisburg, OH, USA).¹¹ The marks produced by this CAD system consisted of 5-cm-diameter circles. For the published results, we used a strict nodule-in-center criterion, which required that the center of a circle was within the lesion boundary for it to be considered a true detection. We did not use a nodule-in-circle criterion, in which the center point of the circle was not required to be within the lesion boundary but the lesion was required only to be located at least partially within the circle, in the previous study¹¹ because we believe that the nodule-in-circle criterion would distort the measurement of performance.

From the Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, MC2026, The University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637, USA.

Correspondence to: Feng Li, Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, MC2026, The University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637, USA; tel: +1-773-8345093; fax: +1-773-7020371; e-mail: feng@uchicago.edu

Copyright © 2009 by Society for Imaging Informatics in Medicine

Online publication 7 May 2009

doi: 10.1007/s10278-009-9201-0

In this study, we analyzed the difference in CAD performance for detection of radiologist-missed cancers on chest radiographs by the FDA-approved version 1.0 that we had used previously¹¹, based on the nodule-in-center criterion and the nodule-in-circle criterion. Recently, Riverain Medical created a new version of their lung nodule detection CAD system, which reduced the number of FP detections by half, while maintaining the same number of true detections. We also compared the performance of this newer CAD version 3.0 with the older version, by applying a recently created database with radiologist-mentioned solitary nodules on chest radiographs. Our purpose in this study was to evaluate the number of actual detections versus "accidental" detections by a CAD system for detection of radiologist-missed lung cancers or radiologist-mentioned solitary lung nodules on chest radiographs, using two different criteria for measuring performance. Detections are considered "accidental" or "detections by chance" when the classification of a lesion as a true positive is based on coincidence due to the shape or size of a marking, rather than the underlying nodule actually being found by the CAD system.

MATERIALS AND METHODS

Institutional review board approval was obtained, and the requirement for informed patient consent was waived. Our study was compliant with the Health Insurance Portability and Accountability Act.

Chest Radiograph Database

Posteroanterior (PA) chest radiographs were obtained with a computed radiography system (Fuji Medical Systems, Stamford, CT, USA) at 110 kVp and 2.5–16 mAs.

The first database that we had used previously¹¹ included 34 patients (21 men and 13 women; mean age of 69 years; age range, 47–87 years) with radiologist-missed cancers on chest radiographs. The records of the cancer registry at the University of Chicago Hospitals were reviewed to identify all patients diagnosed with lung cancer from January 2001 to November 2004 ($n=821$). All patients with chest radiographs on file prior to treatment ($n=314$) were reviewed, and the relevant radiographs

and reports were analyzed. From the 314 patients, 34 PA chest radiographs of 34 patients were identified in which a nodular cancer was present on the PA image in retrospect but had not been mentioned in the report. The location of each radiologist-missed cancer on the 34 chest radiographs was identified by consensus of two radiologists, with all confirmed by computed tomography (CT). All of the cancers were confirmed, using biopsy or surgery as the reference standard.

The second database included 36 patients (17 men and 19 women; mean age of 63 years; age range, 34–98 years). The 36 patients with solitary nodules were selected from the routine clinical workload at the University of Chicago Medical Center from 2005 to 2007. All 36 nodules on the chest radiographs had been mentioned in the original radiology reports. The location of each lung nodule on the 36 chest radiographs was reidentified by two radiologists, and all were confirmed by CT. The final diagnoses for the 36 lung nodules included 21 lung cancers and 15 benign nodules, all of which were confirmed by surgery or follow-up.

The mean diameter (average of the length and width) of each nodule was measured by one radiologist. The lesion boundaries for these nodules were drawn by another radiologist, with the agreement of the first radiologist. Prior to the application of CAD, these nodules were graded for subtlety (from "extremely subtle" to "extremely obvious" on a 1–10 scale) by two radiologists independently, and the numerical ratings by the two were averaged.

CAD System, Sensitivity Criteria

An older FDA-approved nodule detection CAD program (version 1.0, Riverain Medical, Miamisburg, OH, USA) was applied to the 34 chest radiographs with radiologist-missed cancers as well as the 36 chest radiographs with radiologist-mentioned nodules. A newer non-FDA-approved version 3.0 was applied to the 36 chest radiographs with radiologist-mentioned nodules. We did not use the 34 radiologist-missed cancers to test the newer version because some of those cancers had been used to train this newer CAD scheme. The marks applied by these CAD systems consist of 5-cm-diameter circles. The circles were centered about a detection location, which, in the case of a

Fig 1. Two criteria for calculation of CAD sensitivity for nodule detection. (A) A strict nodule-in-center criterion. The center of the circle (*plus sign*) was within the lesion boundary. (B) A generous nodule-in-circle criterion. The center point of the circle (*plus sign*) was not required to be within the lesion boundary, but the lesion was located completely or partially within the circle.

true positive, should correspond to a part of the detected nodule. Therefore, every “genuine” nodule detection should have a part of the nodule at the center of the circle. All CAD marks were analyzed by comparison of the (X , Y) coordinates of the center of each mark to nodule locations determined by the above two radiologists, for assignment of true-positive (TP) and FP results for each chest radiograph.

In this study, two different criteria were compared as methods for computing sensitivity and FP detection rates (Fig. 1). The first was a strict nodule-in-center criterion, which regards a CAD mark as a “true” detection only if the center of the detection circle is located within the lesion boundary. The second criterion was a more

generous nodule-in-circle criterion, which regards a CAD mark as a “true” detection if the lesion boundary is located either partially (at least 40% of the lesion’s area) or completely within the detection circle.

Statistical Analysis

We analyzed the CAD results to determine the number of nodules marked (sensitivity) as well as the number of FP detections. The χ^2 test was performed for the differences in the CAD sensitivities between nodule-in-center criterion and nodule-in-circle criterion for the detection of the 34 radiologist-missed cancers or 36 radiologist-mentioned nodules by use of version 1.0. The χ^2 test was also performed for the differences in the CAD sensitivities between version 3.0 and version 1.0 for the detection of the 36 radiologist-mentioned lung nodules. A P value of 0.05 was used as the threshold for a significant difference.

RESULTS

Among the 34 radiologist-missed cancers (Fig. 2a), 15 were located in the right lung, whereas 19 were located in the left lung. Among the 36 radiologist-mentioned nodules (Fig. 2b), 24 were located in the right lung, whereas only 12 were located in the left lung.

Fig 2. Sites of 34 radiologist-missed cancers and 36 radiologist-mentioned nodules. a Posteroanterior chest radiograph shows 34 radiologist-missed cancers (15 nodules in right lung and 19 in left lung). b Posteroanterior chest radiograph shows 36 radiologist-mentioned nodules (24 nodules in right lung and 12 in left lung).

The mean diameter of the radiologist-missed cancers was 14.5 mm (Table 1). The ratings for cancer subtlety as judged by two radiologists yielded 23 nodules with ratings of 1 to 3 (very subtle), ten with 3.5 to 6, and one with 6.5 to 9 (obvious). The mean diameter of the radiologist-mentioned nodules was 18.6 mm (Table 1). The ratings for cancer subtlety as judged by two radiologists yielded 12 nodules with ratings of 1 to 3, 16 with 3.5 to 6, and eight with 6.5 to 9. No cancers had subtlety ratings greater than 9 for either database.

Table 2 shows CAD (FDA-approved version 1.0) performance with radiologist-missed cancers and radiologist-mentioned nodules on chest radiographs by two criteria. For the radiologist-missed cancers, the sensitivities were 35% (with an average number of 5.9 FP marks per radiograph) and 59% (with 5.6 FP marks), based on the nodule-in-center criterion and the nodule-in-circle criterion, respectively. For the radiologist-mentioned nodules, the sensitivities were 53% (with 5.7 FP marks) and 72% (with 5.5 FP marks), based on the two criteria. No significant differences in the CAD sensitivities between the nodule-in-center criterion and the nodule-in-circle criterion were found for the detection of the 34 radiologist-missed cancers or the 36 radiologist-mentioned nodules (perhaps because of the small database size). The nodule-in-circle criterion caused eight more cancers (24%) to be detected by chance for

Table 1. Subjective Features of Radiologist-Missed Cancers and Radiologist-Mentioned Nodules on Chest Radiographs

Subjective features	Radiologist missed cancers (<i>n</i> = 34)	Radiologist-mentioned nodules (<i>n</i> = 36)
Diameter (mm)		
Mean ± SD	14.5 ± 4.9	18.6 ± 6.4
Range	7–23	8–30
Median	14.8	20.2
Subtlety grade ^a		
1.0–3.0	23	12
3.5–6.0	10	16
6.5–9.0	1	8

All except diameter values are numbers of lung cancers or nodules

SD standard deviation

^aSubtlety was graded (in whole or half [to five tenths] numbers) independently by two observers by using a ten-point scale, with 1 indicating extremely subtle and 10 indicating extremely obvious

Table 2. CAD (Version 1.0) Performance with Radiologist-Missed Cancers and Radiologist-Mentioned Lung Nodules on Chest Radiographs by Two Criteria

CAD performance	Nodule in center	Nodule in circle	<i>P</i> value
Radiologist missed cancers ^a			
Sensitivity	35% (12/34)	59% (20/34)	0.09
False mark rate	5.9 (199/34)	5.6 (191/34)	NA
Radiologist-mentioned nodules ^b			
Sensitivity	53% (19/36)	72% (26/36)	0.14
False mark rate	5.7 (205/36)	5.5 (198/36)	NA

Numbers in parentheses are raw data used to calculate percentages or rates

NA not applicable

^aCAD mark rate per radiograph was 6.2 (211/34) for radiologist-missed cancers, and nodule-in-circle criterion caused eight more cancers (24%) to be detected by chance for the difficult database

^bCAD mark rate per radiograph was 5.9 (224/36) for radiologist-mentioned lung nodules, and nodule-in-circle criterion caused seven more lesions (19%) to be detected by chance

the radiologist-missed cancers and seven more lesions (19%) to be detected by chance for the radiologist-mentioned nodules.

Table 3 shows a comparison of CAD performance with radiologist-mentioned nodules on chest radiographs by two versions. For the 36 nodules, the sensitivities by the version 1.0 system were 53% and 67%, and the sensitivities by the version 3.0 system were 72% and 75%, while reducing the number of FP marks by half (5.7 and

Table 3. Comparison of CAD Performance with Radiologist-Mentioned Nodules on Chest Radiographs by Two Versions

CAD performance	Version 1.0	Version 3.0 ^a	<i>P</i> value
Sensitivity			
Nodule in center	53% (19/36)	67% (24 ^b /36)	0.34
Nodule in circle	72% (26/36)	75% (27 ^b /36)	0.99
False mark rate			
Nodule in center	5.7 (205/36)	2.7 (97/36)	NA
Nodule in circle	5.5 (198/36)	2.6 (94/36)	NA

Numbers in parentheses are raw data used to calculate percentages or rates

NA not applicable

^aVersion 3.0 reduced the number of FP detections by half, while maintaining higher number of true detections by nodule-in-center criterion, and three more lesions (8%) were detected by chance, compared with version 1.0 (19%), by nodule-in-circle criterion

^bThere were 25 or 28 true circles but two marked one nodule

5.5 vs. 2.7 and 2.6), based on the nodule-in-center criterion and the nodule-in-circle criterion, respectively. No significant differences in the CAD sensitivities between version 3.0 and version 1.0 were found for the detection of the 36 radiologist-mentioned nodules. By the version 3.0 system, the nodule-in-circle criterion caused three more lesions (8%) to be detected by chance for the radiologist-mentioned nodules.

Figure 3 shows a cancer for which the center point of the circle was not within the cancer boundary, although the cancer was located partially within the circle. This mark would be counted as a TP by the less strict nodule-in-circle criterion, although the cancer was actually detected by chance.

DISCUSSION

Nishikawa et al.¹ reported that the choice of mammograms used in a database could cause a large variation in sensitivity at four or five FP marks per image and that the sensitivity of their CAD scheme ranged from 26% for a “difficult” database to 100% for an “easy” database at one FP mark per image. Their CAD study for detection of breast masses on mammograms indicated that both the choice of the image database used for training or testing a CAD system and the number of FP marks generated can strongly influence its effectiveness.¹ In a commercially available CAD system for mammograms, the sensitivity has reached higher than 80% for detection of both

breast masses and microcalcifications, with less than 0.3 FP per image.⁹ On a detection CAD system for mammograms, the likelihood of breast lesions being detected by chance has been very low due to the very low number of FP marks. Therefore, the choice of sensitivity measurement criterion may have less impact when the number of false positives is low.

However, the number of FP detections by chest CAD, even by a commercially available CAD system (used in this study), is typically high for detection of lung nodules. Another factor that can greatly influence the calculation of the sensitivity of a CAD system on chest radiographs is a relatively large marking symbol such as the 5-cm circle used in this CAD system. With six to seven CAD marks per image, approximately 25% of the lung area can be covered by the 5-cm circles, greatly increasing the probability that a nodule may be detected by chance.¹¹ Therefore, in order to reduce the number of nodules detected by chance, the choice of criteria for measuring performance used to define true- and false-positive detections could be very important for evaluating the apparent accuracy of a CAD system for detection of nodules on chest radiographs.

A strict nodule-in-center criterion and a generous nodule-in-circle criterion were used for the calculation of the sensitivity of this CAD system. The nodule-in-circle criterion caused eight more cancers (24%) to be detected by chance for a “difficult” database with radiologist-missed can-

Fig 3. Sixty-year-old woman with a T1N0M0 nodular cancer. a Posteroanterior chest radiograph shows a nodular cancer (arrow). b Older version 1.0. The center point of the circle (plus sign) is not within the cancer boundary, but the cancer is located partially within the circle, and this mark was counted as a true positive with seven FP marks by a generous nodule-in-circle criterion. c Newer version 3.0. The center of the circle (plus sign) is within the cancer boundary, and this mark was counted as a true positive with two FP marks only by the stricter nodule-in-center criterion.

cers and seven more lesions (19%) to be detected by chance for a relatively "easy" database with radiologist-mentioned nodules, based on the two criteria. We believe that a substantial number of lesions could be detected by chance with the nodule-in-circle criterion, especially when the testing database includes very subtle lesions such as radiologist-missed cancers and the CAD scheme has a high number of FP detections and/or a large marking symbol.

Certain newer lung nodule detection CAD schemes have improved their sensitivity by incorporating lateral chest images¹² or reducing the number of FP marks by image-processing techniques.¹³ Increasing the sensitivity (including detection of lesions behind the heart or behind the diaphragm) as well as reducing the number of FP marks should be a continuing goal in the development of lung nodule detection systems. In this study, the newer non-FDA-approved version 3.0 of the same CAD system improved the CAD performance of nodule detection; however, it was not the purpose of this study to know the technical reasons for the improvement. We used the radiologist-mentioned nodule database, instead of the original missed cancer database to test the newer version because some of those missed cancers had been used to train this newer CAD scheme and it is imprudent to use the same cases for training and testing of a CAD scheme. For the 36 radiologist-mentioned lung nodules by version 1.0, seven more lesions (19%) were detected by chance and, by version 3.0, three more lesions (8%) were detected by chance, based on the nodule-in-circle measurement criterion. Version 3.0 of the CAD system reduced the number of FP marks by half (two to three marks per image) compared with version 1.0. With three to four CAD marks per image by version 3.0, the area covered by randomly placing circles on the image is decreased compared with version 1.0. These results illustrate that, with a great reduction in the number of FP marks, the likelihood of lesions being detected by chance could also be reduced greatly.

Although reducing the marker size could reduce the number of lesions detected by chance using a nodule-in-circle criterion, we believe that the nodule-in-center criterion is still superior since it helps ensure that the CAD scheme has found the actual lesion and not merely a lesion-like adjoining structure.

CONCLUSION

Our purpose in this study was to evaluate the number of actual detections versus "accidental" detections for lung nodules on chest radiographs by CAD systems using different performance measurement criteria and to show that the choice of performance criteria becomes more critical in systems if they have a larger number of false positives. The specific criteria for measuring performance used to define true- and false-positive CAD detections can substantially influence the apparent accuracy of a CAD system, especially one that has a larger number of false positives.

REFERENCES

1. Nishikawa RM, Giger ML, Doi K, Metz CE, Yin FF, Vyborny CJ, Schmidt RA: Effect of case selection on the performance of computer-aided detection schemes. *Med Phys* 21:265–269, 1994
2. Nishikawa RM, Doi K, Giger ML, Schmidt RA, Vyborny CJ, Monnier-Cholley L, Papaioannou J, Lu P: Computerized detection of clustered microcalcifications: evaluation of performance on mammograms from multiple centers. *RadioGraphics* 15:443–452, 1995
3. Catarious Jr, D, Baydush A, Floyd C: The influence of true positive detection definitions on the performance of a mammographic mass CAD system. *Med Phys* 30:1368, 2003
4. Ikeda DM, Birdwell RL, O'Shaughnessy KF, Sickles EA, Brenner RJ: Computer-aided detection output on 172 subtle findings on normal mammograms previously obtained in women with breast cancer detected at follow-up screening mammography. *Radiology* 230:811–819, 2004
5. Helvie MA, Hadjiiski L, Makariou E, Chan HP, Petrick N, Sahiner B, Lo SCB, Freedman M, Adler D, Bailey J, Blane C, Hoff D, Hunt K, Joyet L, Klein K, Paramagul C, Patterson SK, Roubidoux MA: Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection: pilot clinical trial. *Radiology* 231:208–214, 2004
6. Gar D, Stalder JS, Hardesty LA, Zheng B, Sumkin JH, Chough DM, Shindel BE, Rockette HE: Computer-aided detection performance in mammographic examination of masses: assessment. *Radiology* 233:418–423, 2004
7. Khoo LA, Taylor P, Civen-Wilson RM: Computer-aided detection in the United Kingdom national breast screening programme: prospective study. *Radiology* 237:444–449, 2005
8. Ellis R, Meade A, Mathiason MA, Willison KM, Logan-Logan W: Evaluation of computer-aided detection system in the detection of small invasive breast carcinoma. *Radiology* 245:88–94, 2007
9. Kim JS, Moon WK, Cho N, Cha JH, Kim SM, Im JG: Computer-aided detection in full-field digital mammography: sensitivity and reproducibility in serial examinations. *Radiology* 246:71–80, 2007
10. Kakeda S, Moriya J, Sato H, Aoki T, Watanabe H, Nakata H, Oda N, Katsuragawa S, Yamamoto K, Doi K: Improved

detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system. *AJR* 182:505–510, 2004

11. Li F, Engelmann R, Metz CE, Doi K, MacMahon H: Results obtained by a commercial computer-aided detection (CAD) program with missed lung cancers on chest radiographs. *Radiology* 246:273–280, 2008

12. Shiraishi J, Li F, Doi K: Computer-aided diagnosis for improved detection of lung nodules by use of PA and lateral chest radiographs. *Acad Radiol* 14:28–37, 2007

13. Samei E, Stebbins SA, Dobbins III, JT, Lo JY: Multi-projection correlation imaging for improved detection of pulmonary nodules. *AJR* 188:1239–1245, 2007