Computer-Aided Detection as a Decision Assistant in Chest Radiography

Maurice R.M. Samulski[a], Peter R. Snoeren[a], Bram Platel[b], Bram van Ginneken[a], Laurens Hogeweg[a], Cornelia Schaefer-Prokop[c], and Nico Karssemeijer[a]

[a] Radboud University Nijmegen Medical Centre, The Netherlands
[b] Fraunhofer MeVis, Bremen, Germany
[c] Meander Medical Center Amersfoort, The Netherlands

ABSTRACT

Background. Contrary to what may be expected, finding abnormalities in complex images like pulmonary nodules in chest radiographs is not dominated by time-consuming search strategies but by an almost immediate global interpretation. This was already known in the nineteen-seventies from experiments with briefly flashed chest radiographs. Later on, experiments with eye-trackers showed that abnormalities attracted the attention quite fast but often without further reader actions. Prolonging one's search seldom leads to newly found abnormalities and may even increase the chance of errors. The problem of reading chest radiographs is therefore not dominated by finding the abnormalities, but by interpreting them. **Hypothesis.** This suggests that readers could benefit from computer-aided detection (CAD) systems not so much by their ability to prompt potential abnormalities, but more from their ability to 'interpret' the potential abnormalities. In this paper, this hypothesis was investigated by an observer experiment. **Experiment.** In one condition, the *traditional CAD condition*, the most suspicious CAD locations were shown to the subjects, without telling them the levels of suspiciousness according to CAD. In the other condition, *interactive CAD condition*, levels of suspiciousness were given, but only when readers requested them at specified locations. These two conditions focus on decreasing search errors and decision errors, respectively. Results of reading without CAD were also recorded. Six subjects, all non-radiologists, read 223 chest radiographs in both conditions. CAD results were obtained from the OnGuard 5.0 system developed by Riverain Medical (Miamisburg, Ohio). **Results.** The observer data were analyzed by Location Response Operating Characteristic analysis (LROC). It was found that: 1) With the aid of CAD, the performance is significantly better than without CAD; 2) The performance with interactive CAD is significantly better than with traditional CAD at low false positive rates.

Keywords: Computer-Aided Detection, Chest radiography, Pulmonary Nodules, Observer Study

1. INTRODUCTION

Lung diseases are among the leading causes of death worldwide. Chest radiography is the most common imaging technique for the diagnosis of pulmonary diseases,[1] mainly due to low cost and short examination time. It has been shown that the detection of pulmonary nodules in chest radiographs is extremely difficult, and up to 90% of the cases that were missed contained nodules that were visible in retrospect.[2–4] This shows how important it is to train radiologists in interpreting chest radiographs and developing computer aided detection (CAD) systems to improve detection accuracy.[5,6]

The purpose of the pilot experiment presented in this paper was to investigate how CAD results of chest radiographs can best be presented. Specifically, we investigate two reading methods. In one method, the most suspicious regions according to CAD are shown to the reader on demand without further information on the degree of suspiciousness. In the alternative method, the reader chooses areas of interest and receives on demand the degree of suspicion of that specific area based on CAD analysis. Hence, the first method focuses more on CAD as an aid to **detect** suspicious locations, and the second method focuses more on CAD as an aid to **diagnose** already found suspicious locations.

Further author information: (Send correspondence to Maurice R.M. Samulski)
M.R.M. Samulski, e-mail: m.samulski@rad.umcn.nl, telephone: +31 24 3614766, fax: +31 24 3540866

Medical Imaging 2011: Image Perception, Observer Performance, and Technology Assessment, edited by David J. Manning, Craig K. Abbey, Proc. of SPIE Vol. 7966, 796614 · © 2011 SPIE
CCC code: 1605-7422/11/$18 · doi: 10.1117/12.877968

Figure 1. **Snapshot of the viewing system.** A snapshot made during a session with *interactive CAD prompts*. The label with "1" on the right lung is an annotation made by the reader. The circular region on the left lung is a CAD prompt given after that the subject queried that location with a pointer device. The display provides a CAD computed probability of 0.78 that this region is a TP, while the probability also determines the color of the prompt. After all regions a user wants to report are marked, subjects have to provide ratings of the findings with a slider before they can continue with the next case.

This study is motivated by the following findings. In eye-tracker experiments, Manning et al.[7,8] showed that the error in the detection of pulmonary nodules is mainly explained by decision errors, not search errors; and, with briefly flashed chest radiographs (a few tenths of a second), Kundel and Nodine[9] showed that visual search for pulmonary nodules begins with an almost immediate and global response with surprisingly high performance. Because abnormalities are mostly quickly found but too often wrongly diagnosed we expect that CAD can best be used interactively. This expectation is supported by a similar study in mammography by Samulski et al.[10]

2. METHODS

2.1 Chest Radiographs

The publicly available database of the Scientific Committee of the Japanese Society of Radiological Technology (JSRT, Shiraishi et al.[11]) was used. It consists of 247 posteroanterior chest radiographs; 154 images are abnormal and have ground truth information, and 93 are normal. Each abnormal image contains exactly one pulmonary nodule. The radii of the nodules range from 2.5 mm to almost 30 mm, and the median value is 7.4 mm.

2.2 CAD system

The CAD results of OnGuard$^{\text{TM}}$5.0 (Riverain Medical®, Miamisburg, Ohio) were used. For the JSRT database, OnGuard prompted 386 locations: 105 on normal images and 281 on abnormal images. 43 images have no CAD prompts at all: 31 normal images and 12 abnormal images. For each CAD prompt Riverain Medical provided a computer estimated malignancy score.

A CAD score is in general an abstract measure for suspiciousness that is hardly, or not, to be understood by humans. Therefore, we converted this abstract score into an interpretable measure, namely the probability that the prompted CAD location is inside a truth region, i.e., the probability that it is a true-positive (TP). To prevent bias, the computations were done by a leaving-one-image-out method. The converted CAD score could both be displayed as a number $\in [0, 1]$ and as color coding.

2.3 Workstation

For the purpose of this study, a previously developed workstation is used that has the basic functionality such as zooming, image manipulation, local contrast enhancement and grayscale inversion tools. The brightness and contrast settings were easily adjustable and were set in advance for optimal efficiency. The chest radiographs were viewed on a 30 inch color LCD panel (model FlexScan SX3031W; Eizo Nanao Technologies Inc., Hakui, Ishikawa, Japan) with a native resolution of $2,560 \times 1,600$). On the workstation (see Figure 1) the presence of CAD prompts can be queried interactively by clicking on suspect regions in the chest radiograph using the computer mouse. When a location in the chest radiograph is queried, the workstation checks if a CAD mark exists on that location. If a CAD mark is available, a circle is displayed with the computer-estimated probability. The circle is colored based on the probability of cancer and ranged continously from red to green, for respectively high to low probabilities.

2.4 Experimental Design

The following two conditions are investigated:

A *Interactive CAD.* Areas of interest or suspicion can be sampled interactively by the subject. If a CAD prompt is available at a queried location then it is displayed by a color-coded circle (green: not suspect; red: highly suspect) and a number giving the probability that the prompt is a TP (see Figure 1). Suspect regions are marked by the subject while reading, and scored with ratings between 0 (not suspect) and 100 (highly suspect) after all regions are marked.

B *Traditional CAD.* This condition is split into two sequential parts:

  i *Without CAD.* First, a case is presented without access to CAD and suspect regions are marked. Then the marked regions are scored with ratings between 0 (not suspect) and 100 (highly suspect).

  ii *With CAD.* Next, all available CAD prompts for the image are displayed, but in contrast to the first condition no information is given how suspicious the prompts are. The subject can accept the prompts as relevant and report them, or dismiss them. In addition to adding new findings, the reader has to confirm existing findings reported before CAD prompts were displayed. The subject is also allowed to modify ratings of the annotations previously made without CAD.

The experiment started with a training phase. A representative set of twenty-four chest radiographs were selected from the JSRT database for this purpose. First, subjects were able to see the twenty-four images together with their ground truth and CAD regions. This phase was meant to demonstrate the subtlety of typical pulmonary nodules in the database and to familiarize non-experienced readers with the nodule detection task. Thereafter, the same twenty-four images were presented under the two experimental conditions to become acquainted with the user-interface and to establish a strategy for scoring the suspiciousness of annotations.

The remaining 223 JSRT images were used for the actual experiment. Three random, mutually exclusive sets of images were constructed, consisting of about 75 images each. The sets were stratified by a subtlety rating that is available for all abnormal JSRT images. Each image was seen twice by a subject; once in each of the two experimental conditions. A total of six sessions were done, i.e., two conditions $\times$ three image sets. The order of image sets and reading methods were balanced over the subjects to minimize learning bias. We took care that no images were seen twice in the first three sessions, and only in the last three sessions images are presented for the second time (with the alternative reading method). To decrease a potentially negative effect of remembering cases we demanded a pause of at least a week between the first and the second three sessions.

The subjects' task was to find as many as possible abnormalities with as least as possible mistakes. The subjects were told to treat CAD as an auxiliary tool, a second reader, and not as the leading system. The subjects knew the prevalence of images with pulmonary nodules and they also knew the performance of the CAD system, which was summarized by a FROC curve and the fractions of lesions per probability rating.

2.5 Subjects

Six subjects participated in this experiment. All subjects were non-radiologists involved in radiological research projects. Two of them had extensive experience with reading chest radiographs, two had limited experience, and two had no experience with chest images. All subjects had normal, or corrected to normal, vision and were familiar with the purpose of the experiment.

2.6 Reading times

During the reading sessions reading times per case were automatically recorded. When a subject did not move the mouse and did not do any other action on the workstation for more than 2 minutes this was recorded as idle time in the experiment data file. This idle time is subtracted from the reading time on the basis of the assumption that these excessively long idle times were the result of interruptions during the session. The average reading time per case and its standard deviation was computed for every reader for all three reading modes. Paired reading times were compared by Wilcoxon signed rank testing. A p value of less than 0.05 was considered to indicate a statistically significant difference.

3. RESULTS

The observer data are analyzed by localization receiver operating characteristic analysis (LROC). An annotation or CAD location is considered a true-positive (TP) when it is less than $2\,cm$ from the center of a pulmonary nodule, otherwise it is considered a false-positive (FP). The performances of subjects and the CAD system is given in Figure 2. In clinical routine only few chest radiographs have suspicious nodules. Therefore, the left part of the LROC curves represents the most relevant range. Sensitivity for higher false positive levels is undefined, as readers reported findings in only a fraction of the normal cases. As in previous research,[10] the performance is computed as the mean correct localization fraction in the false-positive fraction interval ranging from 0 to 0.1. With traditional CAD, the performance of the average reader increased at a low false-positive range from 35.2% to 42.8%. When using interactive CAD the performance of the average reader increased from 35.2% to 49.5% in the same false-positive range.

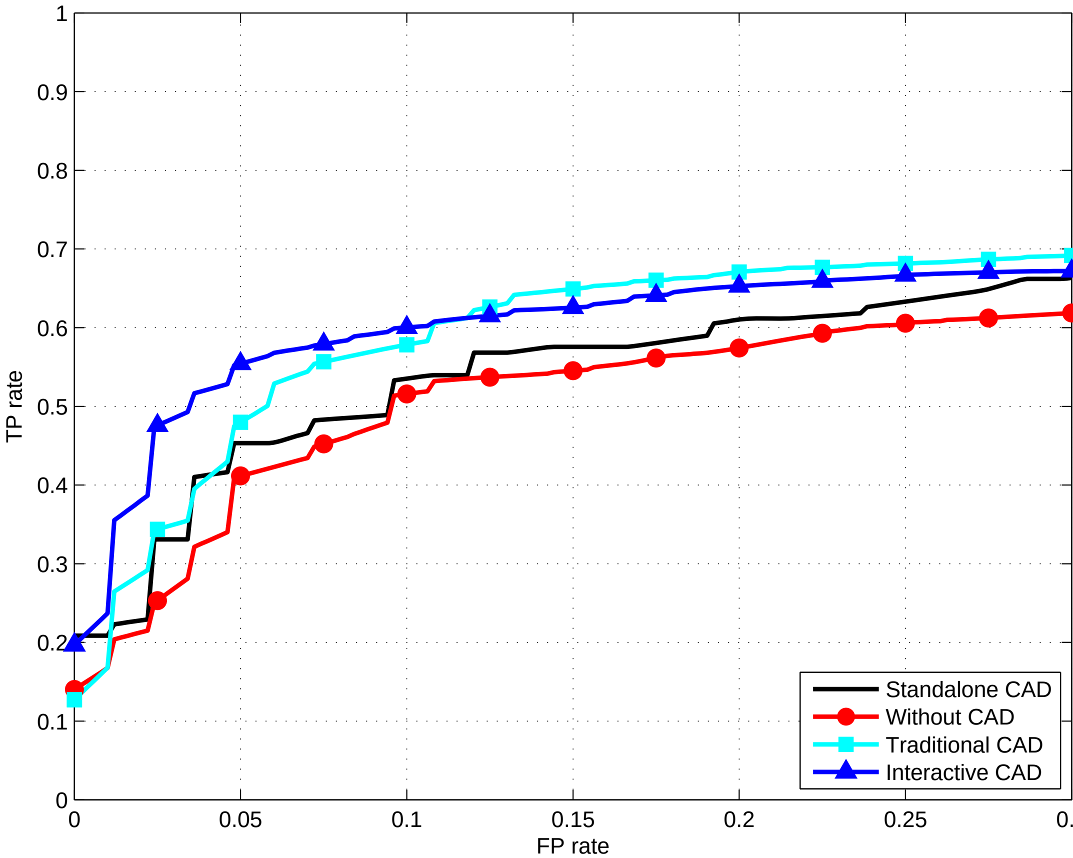
The differences between the columns in the table are significantly different from zero (sign test: $p < .05$).

The average time to read a case without CAD was $34.0 \pm 28.8s$. The average reading time increased to $45.2 \pm 33.8s$ when the traditional CAD prompts were activated and the reader re-evaluated his findings. In the interactive CAD session, the average reading time was $36.5 \pm 28.0s$ (Table 1).

The performance of the CAD system, 39.1% in the false-positive fraction interval 0 to 0.1, appeared to be surprisingly good. CAD standalone was better than the average reader without CAD support.

4. CONCLUSIONS

The performance of the readers with interactive use of CAD is better than with traditional CAD. This confirms the hypothesis that readers take more advantage of CAD as interpretation aid than detection aid. Further improvements might be achievable by adding traditional CAD prompts in the interactive mode, especially if they are rated highly suspicious by CAD and are on regions that were not interactively inspected by the reader, to help overcome occasional perception errors.



(a) Average LROC curves

Subject	Without CAD	Traditional CAD	Interactive CAD
s1	0.286	0.401	0.466
s2	0.394	0.485	0.555
s3	0.376	0.405	0.477
s4	0.427	0.479	0.485
s5	0.302	0.444	0.520
s6	0.324	0.355	0.468
average	0.352	0.428	0.495

(b) Individual Performances

Figure 2. **LROC curves.** Four LROC curves are given in Panel 2(a). Except for the standalone CAD performance, the curves are averages over all subjects. Panel 2(b) gives the individual performances of the subjects. These are computed as the mean correct localization fraction in the false positive fraction ranging from 0 to 0.1. Notably, the standalone detection performance of the CAD system is 0.391.

Moreover, the reading time with interactive CAD is significantly smaller ($p < 0.001$) than reading with traditional CAD. On average, it took only 2.5 seconds longer to read a case using interactive CAD compared to unaided reading of a case.

When lesions are obscured by other tissue, e.g., by the heart, interactive CAD is however inadequate, because lesions are harder to find and thus often not queried. In those cases a hybrid presentation of CAD results may be helpful. In such a system, obscured CAD findings could be presented traditionally and the unobscured CAD findings could be queried interactively.

Subject	Without CAD	Traditional CAD	Interactive CAD
s1	16.7 ± 7.7	28.5 ± 13.6	16.1 ± 6.7
s2	48.2 ± 34.9	61.9 ± 42.9	39.2 ± 24.9
s3	48.6 ± 35.0	57.9 ± 38.3	40.1 ± 26.8
s4	33.6 ± 24.1	47.2 ± 30.3	41.9 ± 25.9
s5	31.9 ± 27.4	42.7 ± 32.5	54.9 ± 38.2
s6	25.3 ± 19.5	33.4 ± 23.4	26.7 ± 16.9
average	34.0 ± 28.8	45.2 ± 33.8	36.5 ± 28.0

Table 1. **Chest radiograph reading times.** Average reading times per case (seconds). Reading times are displayed as mean $\pm$ standard deviation.

ACKNOWLEDGMENTS

The authors would like to thank the readers that participated in the observer study, and Riverain Medical for providing CAD results.

REFERENCES

[1] Mettler, F. A., Bhargavan, M., Faulkner, K., Gilley, D. B., Gray, J. E., Ibbott, G. S., Lipoti, J. A., Mahesh, M., McCrohan, J. L., Stabin, M. G., Thomadsen, B. R., and Yoshizumi, T. T., "Radiologic and nuclear medicine studies in the United States and worldwide: frequency, radiation dose, and comparison with other radiation sources–1950-2007," *Radiology* **253**, 520–531 (2009).

[2] Monnier-Cholley, L., Arrivé, L., Porcel, A., Shehata, K., Dahan, H., Urban, T., Febvre, M., Lebeau, B., and Tubiana, J. M., "Characteristics of missed lung cancer on chest radiographs: a French experience," *European Radiology* **11**(4), 597–605 (2001).

[3] Shah, P. K., Austin, J. H. M., White, C. S., Patel, P., Haramati, L. B., Pearson, G. D. N., Shiau, M. C., and Berkmen, Y. M., "Missed non-small cell lung cancer: Radiographic findings of potentially resectable lesions evident only in retrospect," *Radiology* **226**(1), 235–241 (2003).

[4] Muhm, J. R., Miller, W. E., Fontana, R. S., Sanderson, D. R., and Uhlenhopp, M. A., "Lung cancer detected during a screening program using four-month chest radiographs," *Radiology* **148**, 609–615 (1983).

[5] Abe, H., Macmahon, H., Engelmann, R., Li, Q., Shiraishi, J., Katsuragawa, S., Aoyama, M., Ishida, T., Ashizawa, K., e. Metz, C., and Doi, K., "Computer-aided diagnosis in chest radiography: Results of large-scale observer tests at the 1996–2001 RSNA scientific assemblies," *Radiographics* **23**, 255–265 (2003).

[6] White, C. S., Flukinger, T., Jeudy, J., and Chen, J. J., "Use of a computer-aided detection system to detect missed lung cancer at chest radiography," *Radiology* **252**, 273–281 (2009).

[7] Manning, D., Ethell, S., and Donovan, T., "Categories of observer error from eye tracking and AFROC data," in [*Proceedings of SPIE Medical Imaging*], **5372**, 90–99 (2004).

[8] Manning, D. J., Ethell, S. C., and Donovan, T., "Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph," *The British Journal of Radiology* **77**(915), 231–5 (2004).

[9] Kundel, H. L. and Nodine, C. F., "Interpreting chest radiographs without visual search.," *Radiology* **116**, 527–532 (1975).

[10] Samulski, M., Hupse, R., Boetes, C., Mus, R., den Heeten, G., and Karssemeijer, N., "Using Computer Aided Detection in Mammography as a Decision Support," *European Radiology* **20**, 2323–2330 (2010).

[11] Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., Matsui, M., Fujita, H., Kodera, Y., and Doi, K., "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," *American Journal of Roentgenology* **174**, 71–74 (2000).