



UvA-DARE (Digital Academic Repository)

Advances in digital chest radiography: impact on reader performance

De Boo, D.W.

Publication date
2012

[Link to publication](#)

Citation for published version (APA):

De Boo, D. W. (2012). *Advances in digital chest radiography: impact on reader performance*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Summary
and
General Discussion



Summary

Since the introduction of digital chest radiography continuous advances have been made with respect to detector dose efficiency and processing tools. In this thesis the effect of various aspects of advances in digital chest radiography on actual reader performance was studied. In **Chapter 2** we compared mobile direct radiography (DR) and mobile computed radiography (CR) units for bedside chest radiography of patients admitted to an intensive care unit (ICU). Overall image quality, delineation of anatomical landmarks, and of devices for monitoring were scored better with DR as compared with CR. Even with 50% dose reduction DR outperformed CR with respect to delineation of mediastinal landmarks and devices for monitoring. Image quality was scored equally when assessed individually, in a side-by-side comparison, however, 87% of radiographs obtained with DR at 50% dose reduction were rated superior to CR. Interobserver agreement for the assessment of pathology was used as surrogate to test whether improved image quality would have an effect on diagnostic performance. Only DR achieved an agreement rate of 0.48, which is considered as clinically acceptable. DR_{50%} and CR performed equally (0.39 and 0.33, respectively).

Chapter 3 focuses on the effects of grey-scale reversal for the detection of small pulmonary nodules in chest radiography. Grey-scale reversal is a very simple processing tool, available on PACS workstations with a single mouse click. From optical physiology it is known that optical contrast perception is increased when a dark object is presented on a white background. We suspected that a positive image ("bones black") would help the reader to detect small nodules. Three residents and three radiologists did not benefit from grey-scale reversal for the detection of the nodules (mean diameter 13 mm; median 11 mm) when offered as additional image. We concluded that grey-scale reversal is not a helpful processing tool for the detection of small pulmonary nodules.

Chapter 4 provides an overview of the publications dealing with CAD for the detection of intrapulmonary nodules and T1 lung carcinoma's up to the time when our observer studies were carried out. Very variable results were reported for the various prototypes and FDA approved CAD systems. In all studies, the prevalence of nodules was higher as one would normally see under clinical conditions. Most studies only assessed the stand-alone performance, which is strongly influenced by prevalence, lesion conspicuity and lesion size. Similarly as for the assessment of the stand-alone performance, reader and lesion selection play an important role in observer studies: the level of reader experience and the distribution of lesion conspicuity will strongly

influence the potential effect of CAD. Besides, one CAD algorithm functions differently from the other. All these aspects make it very difficult to compare the results of the various studies with each other and one has to be very careful to draw conclusions on the impact of CAD in general. Experts therefore request to unify evaluation forces and to use “common databases” of clinically validated images.

In **Chapter 5** and **Chapter 6** the two, currently commercially available and FDA approved, CAD systems were tested. Besides their underlying algorithm analysis, they differ in the display of the CAD candidates and the integration into the workflow.

Chapter 5 presents the results of an observer study testing the impact of CAD (Onguard 5.0; Riverain, Miamisburg, Ohio) on the detection of CT proven T1 tumors in participants of the Dutch-Belgian lung cancer screening trial (NELSON). The stand-alone sensitivity of CAD and the sensitivity of the experienced radiologists were equal; however, the mean number of false positives per image was ten times higher for CAD as compared with the radiologists (2.4 vs. 0.24). Use of CAD did not improve the detection performance of the readers though CAD correctly identified tumors, the readers originally had missed. Out of 55 true positive CAD candidates between 5 to 16 were dismissed by the radiologists. Especially true positive CAD candidates for subtle, low conspicuous tumors were dismissed, indicating the potential for CAD, but also the difficulties readers had differentiating true from false positive CAD candidates.

In **Chapter 6** a different patient group and lesion type were selected to test the second, FDA approved, CAD software (IQQA-Chest; EDDA Technology, Princeton Junction, NJ). The study-group compiled of elderly patients, the majority of which had a positive smoking history. Both smoking and aging led to increased interstitial markings described as “dirty lungs”, hampering the detection of focal lesions. Small and low conspicuous pulmonary nodules were included to challenge perception capabilities. Sensitivity significantly increased for inexperienced readers from 39% without CAD to 45% with CAD ($p < 0.05$). A nonsignificant increase of the mean false positives per image (0.27 vs. 0.34) impeded a significant increase of the figure of merit (0.71 vs. 0.71). The experienced readers performed better than the inexperienced, but showed no significant difference in sensitivity (50%) mean false positive per image (0.16 vs. 0.21) or figure of merit (0.84 vs. 0.87) when using CAD as second reader. With 33% of true positive CAD candidates dismissed

and 40% of false positive marks by the readers provoked by a false positive CAD candidate, readers again showed difficulties to differentiate true from false positive CAD candidates.

Based on these results we hypothesized that a lack of “trust” in the performance of CAD obviates a more beneficial use. In **Chapter 7** we therefore studied whether observer training would increase reader performance with CAD for pulmonary nodules. The CAD stand-alone sensitivity was 59%, which was slightly lower than the 65% of the radiologists. Each reader received individual feedback on his / her performance after having read a subset of images. The hypothesis was that this would cause a learning curve, and induce a more beneficial use of CAD. Short-term feedback resulted in an overall increase of reader performance without an improvement by CAD. This was true for both ways CAD can be applied: either with the possibility to discharge marks for lesions located during primary unassisted reading, or for the add-on scenario with preservation of all originally indicated lesions. A total of 32 true positive CAD candidates were dismissed by all readers without a difference over time (16 vs. 16). The dismissed true positive CAD candidates were for low and very low conspicuous nodules in 78% (25/32). These results indicate that short term feedback did not increase the ability of readers to differentiate between true and false positive candidate lesions in order to use CAD more beneficially.

General discussion

Similar as with other imaging techniques (e.g., CT or MRI) the continuous technical development of digital radiography requires constant adaptation of protocols. Scientific evaluation of these advances is necessary in order to estimate the impact on image quality and the value for diagnostic performance. This thesis covers three aspects of recent advances made in the area of digital chest radiography:

- increased dose efficiency for mobile chest radiography
- grey-scale reversal for image display
- computer-aided detection aimed to support the detection of focal pulmonary opacities.

Increased dose efficiency for mobile chest radiography

DR, as opposed to CR, offers higher dose efficiency which had been extensively evaluated in upright chest radiography. Dose reductions up to 60% have been reported⁽¹⁻⁵⁾. Mobile DR units appropriate for bedside chest radiography became technically available only years after those for upright radiography. Until the study, presented in **Chapter 2**, mobile DR units had only been tested for pediatric applications^(6,7). We compared mobile DR and CR units for bedside chest radiography in an adult ICU. One of the major indications for ICU chest radiography, delineation of monitor materials, was scored significantly better with DR as compared with CR, even when DR images were obtained with 50% of the "standard" acquisition dose. The underlying reason is that especially in high absorption areas, such as the mediastinum and upper abdomen, the increased dose efficiency of the DR provides a better signal-to-noise, thus contrast-to-noise ratio. Therefore also the anatomical landmarks within the mediastinum were seen superiorly with DR. Equally for the assessment of overall image quality, DR outperformed CR in the side-to-side comparison. The clinically relevant question is whether this increased image quality can be transferred into higher diagnostic performance. In the absence of a reference standard for pathology, the interobserver agreement was calculated as surrogate for diagnostic performance. Thereby we could demonstrate that the increased image quality of DR indeed transferred into a superior interobserver agreement. DR obtained with 50% dose reduction was still equal to CR. Though the dose of a single radiograph is very low (effective dose: 0.029 mSv), ICU patients frequently undergo multiple follow-up radiographic studies. Whether the increased detector efficiency should be invested into dose reduction for the patient or should be used to increase diagnostic

performance cannot be generally answered and should be determined also as function of the clinical situation and the diagnostic purpose of the examination which is different for upright and bedside radiography. In our institution we chose for patients' dose reduction for bedside radiography.

Grey-scale reversal for image display

Both, clinical experience and the known superior optical contrast perception of the human eye for dark objects on a white background⁽⁸⁾ motivated us to reevaluate grey-scale reversal for the detection of nodular opacities in chest radiography. Gray-scale reversal is a rather simple processing tool and has already been evaluated in the era of hardcopies, more than 20 years ago. At that point, not surprisingly, results had been disappointing, since extra hardcopies with a fixed reversed gradation curve had been produced and tested separately from radiographs with usual gradation characteristics. Nowadays, radiographs are exclusively read as softcopies and gray-scale reversal is available by a simple mouse click on the PACS workstation. The study presented in **Chapter 3** revealed no benefit from grey-scale reversal for the detection of small pulmonary nodules which contradicted our hypothesis. The nodules included in the study had all been rather small with a mean diameter of 13 mm and a median diameter of only 11 mm. We suspected that grey-scale reversal does not help for these types of nodules, due to the interference with vascular and interstitial structures that equally become more prominent with grey-scale reversal. Larger and more ill defined geographic lesions might take more advantage of grey-scale reversal, but this has not yet been tested.

Computer-aided detection aimed to support the detection of focal pulmonary opacities.

Computer-aided detection (CAD) has evolved from prototypes with low sensitivity and high mean false positives per image to FDA approved algorithms with reported stand-alone sensitivities of 35% and 47% for initially missed bronchogenic carcinomas^(9,10). A lower agreement rate between CAD and radiologists compared to the agreement between radiologists furthermore underscores the potential of CAD to detect nodules that radiologists tend to miss⁽¹¹⁾. The literature lists some publications that report a promising increase of reader performance with the use of CAD as second reader⁽¹²⁻¹⁴⁾. These results are in contrast to the data presented in **Chapter 5** and **Chapter 6**. Both studies failed to show a significantly improved detection performance of observers for small focal lung lesions. Nevertheless we observed an interaction between observers and CAD.

In the first study 82% (54/66) of new markings made by observers were due to false positive CAD candidates. In the second study 40% (61/154) of detrimental rating differences were possibly provoked by a false positive CAD candidate in the same location. True positive CAD candidates were dismissed in 80% and 33% respectively, meaning the observer interpreted the region pointed out by CAD as normal. The majority of these dismissed true positive candidates were for low to very low conspicuous lesions. We concluded from these results that CAD has potential to have a positive impact on reader performance since it detects different lesions than the observers. A more beneficial use of CAD was undone by the inability of the observers to discriminate true positive from false positive CAD candidates. It is likely that high and moderate conspicuous lesions that have been missed by the readers due to "inattentive blindness" would take more easily advantage of the availability of CAD. This effect, however, is very difficult to prove under study conditions, since observers will always analyze images with special diligence under study conditions. Lesions of low or very low conspicuity, however, have different diagnostic requirements; correct diagnosis requires not only visual localization but also correct differentiation from surrounding "anatomic noise". None of the observers in our study had a broader clinical experience with the CAD algorithm. Though we had introduced them to the software by a number of training cases, we suspected that the inability to correctly discriminate the candidates could have been also caused by a lack of familiarity with the system. We therefore undertook the study evaluating the effects of short term feedback to the readers described in **Chapter 7**. All readers received an individual feedback on their performance without and with CAD immediately after having read a subset of the study images. Results, however, found no increase of readers' ability to differentiate true from false positive CAD candidates. There was an overall training effect for the detection of nodules but not an increased benefit of CAD. The ability to discriminate true from false positive candidates is apparently more complex and requires more information than increasing the familiarity with the CAD algorithm itself. Limited evidence is available about observer training for CAD. A one-day-training for CAD in CT colonography resulted in an increase of sensitivity, but also in a decrease of specificity⁽¹⁵⁾. For CAD applied in mammography an increase of sensitivity, specificity and Az was reported after a four week training period⁽¹⁶⁾. However, the ultimate learning curve for CAD in mammography was estimated to be around 2 years⁽¹⁷⁾. Whether this applies for CAD in chest radiography is not yet known. Further on it is very likely that an increase of confidence into CAD can be achieved by decreasing the number of false positive CAD candidates. Li et al. reported that 68% of

the false positives CAD candidates were solely or in part triggered by a bony structure⁽⁹⁾. First studies combining CAD with bone suppression techniques, like energy subtraction, provided promising results^(18,19). Studies under way using an upgraded software version of the same CAD algorithm used in chapter 6 (Onguard; Riverain), showed a marked performance improvement with a sensitivity of around 70-80% with a mean number of false positives per image of 0.5. Whether this indeed transfers to an improved observer performance is currently under evaluation.

Another approach is to alter the presentation of the CAD candidates: currently a maximum of 5 ROIs are presented on demand without any additional information. An option could be to add the likelihood per CAD candidate. Another option could be to leave the visual interrogation of the chest radiograph to the observer and only provide CAD results for locations the observer is pointing out. The latter approach was found successful for the detection of tumors in mammography⁽²⁰⁾ and produced promising results for nodule detection in a preliminary study using a public data base⁽²¹⁾ and a number of observers not trained in radiology⁽²²⁾.

All of these above mentioned approaches are currently under evaluation and upcoming results will determine the future role of CAD for nodule detection in chest radiography.

In summary this thesis demonstrates that

- The introduction of mobile direct radiography at the bedside allows for 50% dose reduction, as compared to computed radiography, without loss of clinically relevant image quality. Alternatively, the improved image quality obtained at unaltered dose can be used to uniform diagnostic performance.
- Using PACS display of digital chest radiographs, gray-scale reversal does not help the radiologists in detecting small pulmonary nodules.
- The potential of CAD to reduce detection errors by radiologists is not fully established.
- Despite short-term observer training, radiologists still have difficulties differentiating true positive from false positive CAD candidates.

References

- 1 Gruber M, Uffmann M, Weber M, Prokop M, Balassy C, Schaefer-Prokop C. Direct detector radiography versus dual reading computed radiography: feasibility of dose reduction in chest radiography. *Eur Radiol* 2006; 16:1544-1550
- 2 Strotzer M, Volk M, Frund R, Hamer O, Zorger N, Feuerbach S. Routine chest radiography using a flat-panel detector: image quality at standard detector dose and 33% dose reduction. *AJR Am J Roentgenol* 2002; 178:169-171
- 3 Herrmann A, Bonel H, Stabler A, et al. Chest imaging with flat-panel detector at low and standard doses: comparison with storage phosphor technology in normal patients. *Eur Radiol* 2002; 12:385-390
- 4 Bacher K, Smeets P, Bonnarens K, De Hauwere A, Verstraete K, Thierens H. Dose reduction in patients undergoing chest imaging: digital amorphous silicon flat-panel detector radiography versus conventional film-screen radiography and phosphor-based computed radiography. *AJR Am J Roentgenol* 2003; 181:923-929
- 5 Bacher K, Smeets p, Vereecken L, et al. Image quality and radiation dose on digital chest imaging: comparison of amorphous silicon and amorphous selenium flat-panel systems. *AJR Am J Roentgenol* 2006; 187:630-637
- 6 Rapp-Bernhardt U, Bernhardt TM, Lenzen H, et al. Experimental evaluation of a portable indirect flat-panel detector for the pediatric chest comparison with storage phosphor radiography at different exposures by using a chest phantom. *Radiology* 2005; 237:485-491
- 7 Rapp-Bernhardt U, Roehl FW, Esseling R, et al. Portable flat-panel detector for low-dose imaging in a pediatric intensive care unit comparison with an asymmetric film-screen system. *Invest Radiol* 2005; 40:736-741
- 8 Blackwell H. Contrast thresholds of the human eye. *J Opt Soc Am* 1946; 36:624-643
- 9 Li F, Engelmann R, Metz CE, Doi K, MacMahon H. Lung cancers missed on chest radiographs: results obtained with a commercial computer-aided detection program. *Radiology* 2008; 246:273-280
- 10 White CS, Flukinger T, Jeudy J, Chen JJ. Use of a computer-aided detection system to detect missed lung cancer at chest radiography. *Radiology* 2009; 252:273-281
- 11 Bley TA, Baumann T, Saueressig U et al. Comparison of radiologist and CAD performance in the detection of CT-confirmed subtle pulmonary nodules on digital chest radiographs. *Invest Radiol* 2008; 43:343-348
- 12 Van Beek EJ, Mullan B, Thompson B. Evaluation of a real-time interactive pulmonary nodule analysis system on chest digital radiographic images: a prospective study. *Acad Radiol* 2008; 15:571-575
- 13 Xu Y, Ma D, He W. Assessing the use of digital radiography and a real-time interactive pulmonary nodule analysis system for large population lung cancer screening. *Eur J Radiol* 2011 May 27 [Epub ahead of print]
- 14 Song W, Fan L, Xie Y, Qian JZ, Jin Z. A study of inter-observer variations of pulmonary nodule marking and characterizing on DR images. *Proc SPIE* 2008; 5749:272-280
- 15 Taylor SA, Burling D, Roddie M, Heneyfield L, McQuillan J, Bassett P, Halligan S. Computer-aided detection for CT colonography: incremental benefit of observer training. *Br J Radiol* 2008; 81:180-186
- 16 Luo P, Qian W, Romilly P. CAD-Aided mammogram training. *Acad Rad* 2005; 12:1039-1048.
- 17 Nishikawa R. Increased CAD use prompts look at advantages, drawbacks. *Radiological Society of North America Web site*. <http://www.rsna.org/Publications/>

- rsnanews/feb07/upload/RSNANews_Feb07_CAD_Usage.pdf. Published February 2007. Last accessed September 18, 2011
- 18 Szucs-Farkas Z, Patak MA, Yuksel-Hatz S, Ruder T, Vock P. Improved detection of pulmonary nodules on energy-subtracted chest radiographs with a commercial computer-aided diagnosis software: comparison with human observers. *Eur Radiol* 2010; 20:1289-1296
 - 19 Balkman JD, Mehandru S, DuPont E, Novak RD, Gilkeson RC. Dual energy subtraction digital radiography improves performance of a next generation computer-aided detection program. *J Thoracic imaging* 2010; 25:41-47
 - 20 Samulski M, Hupse R, Boetes C, Mus RD, den Heeten GJ, Karssemeijer N (2010) Using computer-aided detection in mammography as a decision support. *Eur Radiol* 20:2323-2330
 - 21 Shiraishi J, Katsuragawa S, Ikezoe J, et al. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *AJR Am J Roentgenol* 2000; 174:71-74
 - 22 Samulski MRM, Snoeren PR, Platel B, et al. Computer-aided detection as a decision assistant in chest radiography. *Proc. SPIE* 2011; 796614