



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.

Computer-Aided Detection (CAD) of Lung Nodules and Small Tumours on Chest Radiographs



Diederick W De Boo
Mathias M Prokop
Martin Uffmann
Bram van Ginneken
Cornelia M Schaefer-Prokop

European Journal of Radiology 2009; 72:218-225

Abstract

Detection of focal pulmonary lesions is limited by quantum and anatomic noise and highly influenced by variable perception capacity of the reader. Multiple studies have proven that lesions - missed at time of primary interpretation - were visible on the chest radiographs in retrospect. Computer-aided detection (CAD) schemes do not alter the anatomic noise but aim to decrease the intrinsic limitations and variations of human perception by alerting the reader to suspicious areas in a chest radiograph when used as a "second reader". Multiple studies have shown that the detection performance can be improved using CAD, especially for less experienced readers, at a variable amount of loss of specificity. There seems to be a substantial learning process for both experienced and inexperienced readers, to be able to optimally differentiate between false positive and true positive candidates. Readers have to build up sufficient "trust" in the capabilities of these systems to be able to use them at their full advantage. Studies so far focussed on stand-alone performance of the CAD schemes to reveal the magnitude of potential impact, or on retrospective evaluation of CAD as second reader for selected study groups. Further research is needed to assess the performance of these systems in clinical routine and to determine the trade-off between performance increase in terms of increased sensitivity, decreased interobserver variability, loss of specificity, and secondary indicated follow-up examinations for further diagnostic work up.

Background

Lung cancer is the second most commonly diagnosed cancer and the leading cause of cancer-related deaths in the United States. Despite increasing awareness of the deleterious effects of smoking and continuous research in lung cancer diagnosis and treatment, lung cancer mortality is very high and has not changed substantially over the last decade. Yet, it is proven that bronchogenic carcinoma has a much better chance of cure if diagnosed at an early (localized) stage⁽¹⁾. Lung cancer screening using CT is presently subject to intense research with a number of randomized and non-randomized trials still ongoing. While CT is very sensitive for detection of small pulmonary lesions, the vast majority of detected nodules are in fact benign⁽²⁾. The verdict is not yet out whether CT screening is reducing lung cancer-related mortality or in fact leads to a large number of unnecessary interventions and ultimately no survival benefit. Lung cancer screening efforts in the 1980s had focused on the use of chest radiographs. While more cancers were detected and more resections were performed, there was no survival benefit for screened participants. Since then the technique of chest radiography has seen vast improvements including the transition to digital radiography with better contrast resolution and better depiction of previously difficult areas, such as the lung recesses and the retrocardiac space. Detection performance for small lung tumours should therefore be improved. Whether this translates into better patient survival, however, remains to be seen. Lung cancers missed on chest radiographs, however, remain one of the major reasons for lawsuits against radiologists⁽³⁾. Investigators of the Mayo Lung Project reported 75% of the perihilar (12/16) and 90% of peripheral nodules (45/50) had been missed at time of primary interpretation, but were visible on the chest radiographs in retrospect⁽⁴⁾. Similar results were reported by other publications^(5,6). For this reason, detection of suspicious nodules remains a main task for any radiologist reading chest radiographs, even if the primary diagnostic question is not cancer-related. Though its inferior sensitivity to CT, chest radiography represents a fast and relatively cheap imaging modality. It remains a popular modality for the surveillance of pulmonary metastatic disease in patients with known malignancies. Also, in most cases it is the primary method of choice to screen for intrapulmonary abnormalities.

Rationale for CAD

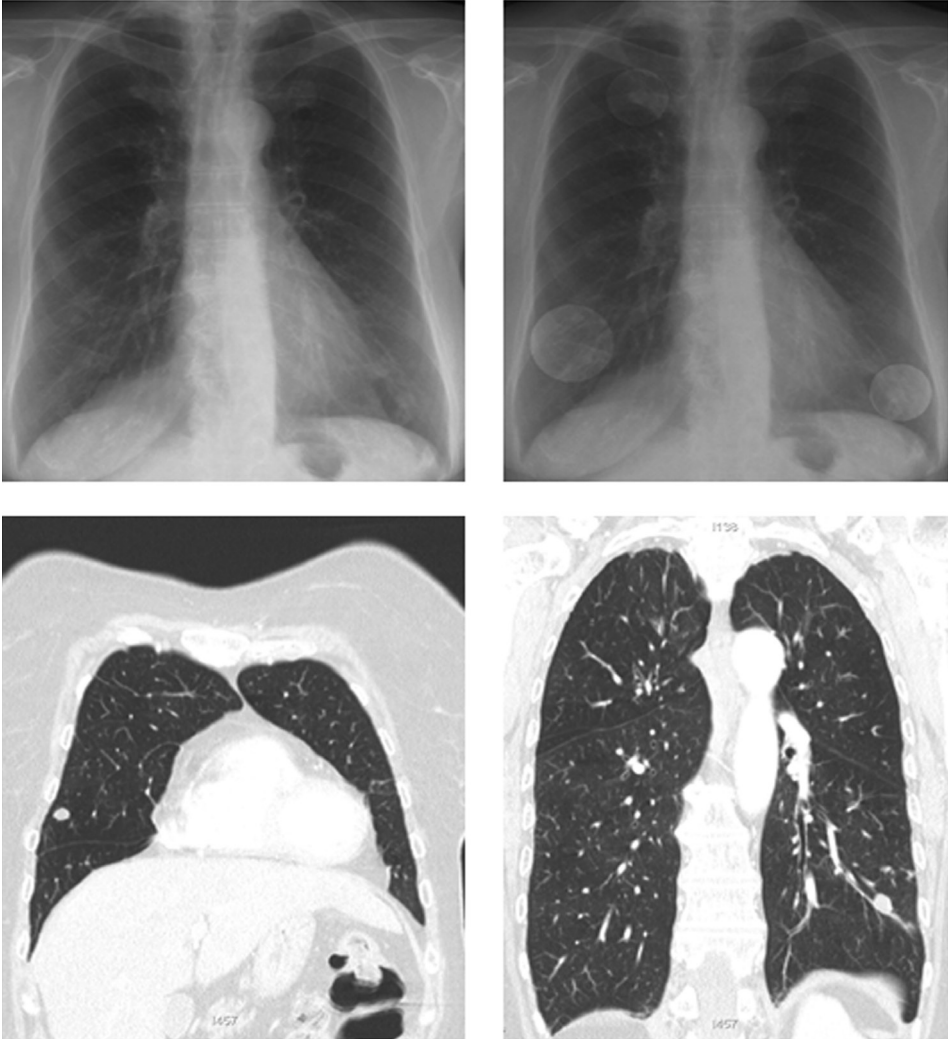
Detection of subtle focal pulmonary opacities on radiographs remains a challenge. It is limited by two categories of noise: 1) the radiographic noise (mottle) related to the quantum nature of radiation^(7,8) and 2) the anatomic noise⁽⁹⁾ which refers to surrounding and overprojecting anatomic structures, such as ribs, abnormalities in the lung parenchyma and vascular structures. The term “conspicuity” of a lesion describes the relation of feature contrast to surrounding complexity, thus including the contributions of both noise categories⁽¹⁰⁾. In chest radiography, anatomic noise appears to have far greater influence on the detection of pulmonary nodules^(11,12). The complexity of the surrounding anatomic noise greatly influences the perception of the radiologists. In an experiment, in which the authors slightly varied the location of simulated nodules, Samei et al. found strong correlations between nodule size, nodule location and observer detection performance⁽⁹⁾. Anatomic structures such as ribs and pulmonary vessels superimposing on a subtle lung nodule on a chest radiograph influenced nodule detection, measured as the area under the receiver operating characteristics (ROC) curve (A_z), by as much as 28%. The effects of distracting anatomic noise are aggravated by the intrinsic limitations of human perception. Perception is influenced by predictable measures such as training and experience, but also by less predictable effects such as concentration, distraction and fatigue. Radiologists are not consistent in what they detect and what they diagnose. The less conspicuous the findings, the greater the variability will be. When confronted with a set of radiographs a second time, the reader may detect different lesions than during the first time. Studies have repeatedly shown that missed lesions were evident in retrospect⁽⁴⁻⁶⁾. To improve interpretation in chest radiography, various technical innovations have been introduced, among which image processing techniques represent the most important tools. Image processing includes techniques such as dual energy subtraction, temporal subtraction and tomosynthesis all aim to increase the conspicuity of a lesion by reducing the impact of anatomic noise. Computer-aided detection (CAD) algorithms have a different approach; they do not alter the anatomic noise. By alerting the reader to suspicious areas in the chest radiograph CAD aims to decrease the intrinsic limitations and variations of human perception. In addition to prototypes only applied under research conditions, there are currently two US Food and Drug Administration (FDA)-approved systems available on the market (IQQA-chest, EDDA technology, Inc Princeton Junction, NJ, USA; Figure 1) and ONGUARD; Riverain Medical, Miamisburg, Ohio, USA; Figure 2), more systems from other manufacturers are expected to follow.

For mammography, the first FDA approved clinical system was introduced already in 1998. Since then systems were continuously improved and received a widespread adoption in the USA, where there is additional reimbursement for the use of CAD⁽¹³⁾. Though being in use already for a longer time, the clinical value of CAD is still being debated⁽¹⁴⁾.

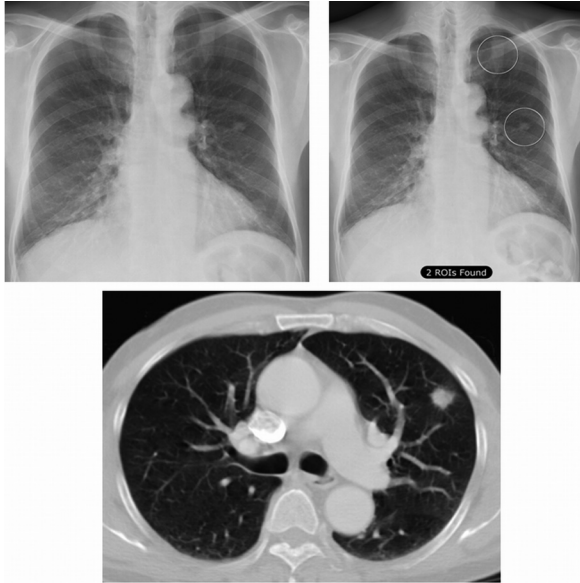
Prerequisites for routine implementation of CAD

Computerized methods for automatic detection of lung nodules were published already in the 1970s. None of these methods, however, were applied in clinical studies at this point, probably because of large numbers of false positive findings generated with these early methods⁽¹⁵⁾. A study from 1993 postulated the potential usefulness of CAD schemes for nodule detection if the false positive rate could be reduced to a level of approximately one (!) false positive detection per radiograph at a sensitivity of 75% ⁽¹⁶⁾. More than 15 years later, even the most advanced CAD algorithms do not live up to this postulated requirement. In the meantime CAD has become clinically available for mammography, and it now appears to be ready also for broader clinical application in the chest. The reason for this difference in acceptance between CAD for mammography and CAD for chest radiography may be based on the different tasks. In mammography the detection task is two-fold and refers to the localisation of small, but high-contrast microcalcifications on one side and the detection of ill-defined soft-tissue masses on the other side. While for the first, the sensitivity of modern CAD systems is reported to be very high (98%), the sensitivity for mammographic masses is significant lower⁽¹⁷⁾. Pulmonary nodules and masses, however, seem to face similar problems with respect to conspicuity and their vulnerability towards the effect of overlying and distracting anatomic structures. All CAD systems are designed to be used as a second reader: they are meant as complementary tool that draws the radiologists' attention to certain image areas that need further evaluation. They are not designed to detect all potential lesions, which would allow the radiologist only to focus on the areas identified by the CAD system. The present systems do not provide a 100% sensitivity, which makes it necessary for the radiologist still to evaluate the whole image. However, the systems can detect additional lesions that might escape the radiologist's attention. In clinical practice, the CAD program runs in the background while the radiologist completes visual interrogation of the posteroanterior (PA) and lateral radiograph. Subsequently suspicious lesions detected by the CAD are revealed and the radiologist has the opportunity to accept or discard the CAD findings.

Figure 1



Two intrapulmonary metastases (see CT), both correctly assigned by the CAD (IQQA-chest, EDDA).

Figure 2

Correctly assigned CT proven T1 tumor (ONGUARD, Riverain) in the left upper lobe and false positive candidate in the left long apex due to crossing of clavicle and first rib.

Usually the findings are indicated by a region of interest (ROI) around the suspicious lesion or by highlighting the suspicious areas. Comparing the original image with CAD candidates can be done side-by-side or by toggling the highlighted areas / circles on and off. Whatever implementation, the application of CAD results in additional reading effort and increased reading time at least for a substantial number of cases, certain readers, and within an initial time period of adaptation. Whether reading time will be generally increased or only before adaptation to the system remains to be evaluated. In order to make these additional investments worth their while, the number of false positive candidates has to be low and the true positive candidate lesions should complement the lesions found by the radiologists: there is no advantage if the lesions found by CAD completely overlap the lesions found by human observers. Another prerequisite for clinical practice is the seamless integration of the CAD into a picture archiving and communication system (PACS). CAD results have to be available on-demand and need to be presented in a fashion that disturbs the normal workflow as little as possible. In addition, CAD should work independently of which digital imaging system has been used for the acquisition of the digital chest radiographs. A few logistic and potentially medico-legal issues have not yet been resolved; whether and for how long candidate lesions have to be stored, and whether CAD candidates should be visible only for radiologists or also for the referring physicians⁽¹⁸⁾. CAD will most inevitably detect lesions that turn out to be true tumours but have been refuted by the

radiologist at time of the evaluation. This may lead to more defensive medicine with unnecessary follow-up radiographs or CT examinations if lawyers are allowed to interpret such a situation as an instance of malpractice. In addition, it will be difficult for less experienced clinical colleagues to differentiate between true and false positives candidates. Again this holds potential for improper follow-ups and conflict between radiologists and referring physicians. For all these reasons, it is probably most appropriate not to store the CAD results in a PACS environment.

CAD results as published so far

Study designs

There are various study setups and statistical models that have been applied to assess the effects of CAD (Table 1). Selection and prevalence of lesions, number and experience of readers, standard of truth, and statistical analysis, represent factors that influence the results and should be considered when drawing conclusions and comparing performance of various CAD systems. The majority of studies are based on a set of radiographs that was selected to include a certain number of exams containing nodules or tumours and a certain number of controls (Table 1). Ideally, the presence or absence of lesions is proven by CT as the superior gold standard. However, such a process introduces a selection bias that depends on the selected type of abnormality and does not reflect the normal clinical situation in which there is no superior standard available for most patients. CAD performance has two components, the stand-alone performance of CAD without human interaction and the effect of CAD on reader performance. Stand-alone performance is a good indicator of the magnitude of the potential effects of CAD; it determines how many and what kind of lesions can be detected and how many false positives per image are produced. Sensitivity and false positives are interrelated: higher sensitivity for a specific CAD algorithm always comes with a higher number of false positive CAD candidates per image. For this reason a compromise has to be made between these two factors. In some studies the operating point was variable and allowed for analysis of this interrelation, in other studies the operating point is fixed and provides only one set of sensitivity and number of false positive CAD candidates per image. Because CAD algorithms cannot detect all potential lesions in a radiograph at a reasonable number false positive candidates, they need to be used as second reader. How much a CAD algorithm is able to improve the performance of a reader depends on the following factors:

- The reader experience; less experience will lead to a potentially greater increase in performance.
- The stand-alone performance of CAD; the higher the detection rate of nodules that are typically missed by human observers, the more effect a CAD will have.
- The ability of observers to distinguish between a true positive and a false positive CAD candidates. Especially the latter has not yet been extensively studied, but is an important factor that will ultimately determine whether CAD will mainly increase the numbers of true lesions found by an observer or whether it will also increase the number of false positives by the observer. It is therefore important that readers become familiar with the behaviour of a "their" CAD system so that they learn the optimum cut-off for differentiating true and false positive candidates. As a result of the unfamiliarity with CAD, observers may dismiss true positive candidates or lose confidence in CAD, because of too many false positive candidates. Statistical analysis of data is important for interpreting the outcome of CAD studies; sensitivity and specificity are influenced by the prevalence of disease within the study group, especially if a high prevalence is known or suspected by the readers. ROC incorporates the interrelation of sensitivity and specificity of human observers depending on their confidence level. Therefore, ROC analysis are to be preferred. Besides that, ROC statistics work on a per-region basis; any positive reading of a region containing a true lesion is counted as true positive, independent of whether the reader had actually identified the lesion or had read a false positive contained in the same region. Localized ROC (L-ROC) analysis forces the reader to indicate a lesion and thus also incorporates the correct localization of a lesion. Any type of ROC statistics, however, requires the availability of a superior gold standard, usually provided by CT examinations.

CAD for nodule detection

First publications⁽¹⁹⁾ are more than 10 years old and were based on the comparison of conventional radiographs and digitised versions of these radiographs that included the superimposed CAD output. Although it can be expected that CAD software has further improved in the meantime, already this first study reported excellent results; diagnostic accuracy improved significantly with CAD (ROC area 0.906 vs. 0.948) while reading time did not increase significantly. The authors also reported a larger benefit for inexperienced radiologists using CAD and demonstrated a subsequent decrease in variability of accuracy across readers of different experience levels. The first study that brought the potential of CAD for nodule detection

to the attention of a large group of radiologists was a large-scale observer test conducted during the 1996 RSNA scientific assembly in Chicago⁽²⁰⁾. Radiologists at the RSNA were invited to evaluate 22 abnormal and 20 normal digitised chest radiographs in random order without and then with CAD. The CAD algorithm had a stand-alone performance of 70% with a mean of only one false positive CAD candidate per radiograph. The 146 observers included chest radiologists, general radiologists and residents. For all categories of readers an improved detection rate was seen with the application of CAD. Az varied from 0.697 to 0.825 without CAD and from 0.80 to 0.88 with CAD. The correct diagnosis was shown after each case, which is unrealistic under clinical conditions, but is likely to have introduced a learning effect. Learning how to use CAD, is crucial. Understanding which lesions CAD is likely to detect and which it is likely to miss is important for a reader to make the decision whether to accept or dismiss a CAD candidate. The reduction in variability between radiologists of different experience was further confirmed by other studies; the less experienced the observer, the greater the improvement in lung nodule detection with CAD will be⁽²¹⁻²³⁾. In 2008 Bley et al. published the stand-alone performance of one of the first commercial CAD systems (IQQA-chest, EDDA technology) for the detection of CT-proven nodules with a mean diameter of 7.5 mm \pm 2.2 mm⁽²⁴⁾. CAD yielded a sensitivity of 39% (stand-alone performance) compared to sensitivities between 18% and 30% for the radiologists. Most interestingly, the agreement among radiologists was larger ($\kappa = 0.64-0.73$) than between radiologists and the CAD algorithm ($\kappa = 0.45-0.52$). This indicates that CAD indeed detected different lesions than radiologists. An important step towards an increased comparability of the performance of various CAD algorithms would be their evaluation using the same study cases; Schilham et al. assessed their own CAD algorithm developed in academia using the Japanese Society of Radiological Technology (JSRT) database, a large and publicly available database. They reported a stand-alone sensitivity of 51% with 2 false positive CAD candidates per image, and a stand-alone sensitivity of 67% with 4 false positive CAD candidates per image⁽²⁵⁾. Other authors detected 78% of the nodules, but at 4 false positive CAD candidates per image⁽²⁶⁾. These results nicely demonstrate the inevitable trade-off between sensitivity and specificity. Kasai et al. evaluated the effect of CAD on the detection of subtle nodules in PA and lateral chest radiographs⁽²⁷⁾. About 50% of the nodules (15/31) were visible only on the PA views. The CAD software indicated candidate lesions on both views. The authors found an increased sensitivity (65% vs. 68%) with CAD, but no significant increase of the Az (0.804 vs. 0.816, $p = 0.297$) indicating that the increased sensitivity was

counteracted by an increased number of false positive marks with CAD. The only study so far that tried to assess the usefulness of CAD in a clinical environment was published in 2008 by van Beek et al⁽²⁸⁾. They investigated the benefit of CAD for nodule detection on chest radiographs obtained in daily practice. The study group exclusively consisted of patients with known extra-pulmonary primary malignancies, who were under surveillance for metastatic disease. In 214 of the 324 chest radiographs a standard of reference could be established by CT or follow-up of at least 6 months. Nodules were present on 55 of the 214 chest radiographs. The sensitivity significantly increased with CAD from 64% to 93%; with only a nonsignificant decrease in specificity (93% vs. 96%).

Table 1

List of publications using FDA approved and prototypes of CAD schemes for the detection of nodules and T1 bronchogenic carcinomas.

Author	System	n	Prevalence	CAD stand-alone	FP per image by CAD	Reader vs. reader with CAD
CAD for pulmonary nodules						
v Beek (2008)	EDDA	214	26%	-	-	sensitivity 64% vs. 93%*
Bley (2008)	EDDA	117	36%	39%	2.7	-
Hardie (2008)	not FDA	154	100%	78%	4	-
He (2008)	EDDA	116	50%	67%	2.4	-
Kasai (2008)	not FDA	60	52%	52%	4.2	sensitivity 65% vs. 68%* Az 0.804 vs. 0.816
Shiraishi (2007)	not FDA	106	100%	71%	4.9	-
Schillham (2006)	not FDA	247	62%	51%	2	-
				67%	4	-
Shiraishi (2006)	not FDA	48	75%	62%	-	Az 0.724 vs. 0.778*
Shiraishi (2006)	Riverain	459	100%	70%	5	
Song (2005)	EDDA	232	36%	71%	2.8	sensitivity 49% vs. 80%*
Coppini (2003)	Riverain	128	-	60% **	4,3	
				75% **	10.2	
Shiraishi (2003)	not FDA	90	60%	100% **	3.1	Az 0.682 vs. 0.808*
Freedman (2002)	not FDA	240	67%	66%	5	Az 0.835 vs. 0.865*
MacMahon (1999)	not FDA	40	50%	80% **	1 **	Az 0,825 vs. 0.889*
Xu (1997)	not FDA	200	50%	70%	1.7	-
Kobayashi (1996)	not FDA	120	50%	-	-	Independent Az 0.894 vs 0.940*
Matsumoto (1992)	not FDA	198	48%	60%	15	-
CAD for T1 lungcarcinoma						
White (2009)	Riverain	114	100%	47%	3.9	-
Li (2008)	Riverain	34	100%	34%	5.9	-
Sakai (2006)	not FDA	100	50%	74%	2.2	Az 0.896 vs. 0.923*
Kakeda (2004)	not FDA	90	50%	-	3.2	Az 0.924 vs. 0.986*

FP: false positives; Az: area under ROC curve.

* significant difference

** fixed

In summary, CAD for nodule detection has shown to increase reader performance with better sensitivity, especially of inexperienced readers, and to decrease inter- and intra-reader variability.

CAD for detection of early stage bronchogenic cancer

The first studies evaluating the effects of CAD on the detection of T1 tumours were published by Japanese colleagues. This task may differ from detecting metastases because nodule characteristics of T1 tumours are often different from those of metastases. A significantly increased detection rate for early lung tumours ($n = 45$, mean diameter 18 mm; range between 8 and 25 mm) was described by Kakeda et al. who evaluated CAD using 8 radiologists of varying experience⁽²¹⁾. The Az increased from 0.924 to 0.986, and thus started already at a very high baseline. Board certified radiologists performed better and profited less from CAD than residents, but even residents had a baseline performance without CAD that exceeded 0.90. These results were confirmed by Sakai et al. who also assessed the detection of T1 tumours by 8 readers of varying experience⁽²⁹⁾. Four chest radiologists and 4 residents evaluated PA radiographs of 50 patients with T1 tumours (< 3 cm in diameter) and 50 controls. Stand-alone performance of CAD included a sensitivity of 74% and a mean number of false positive CAD candidates of 2.3. Az significantly increased from 0.896 to 0.924 with CAD. CAD might help to reduce the number of T1 bronchogenic tumours originally missed in the reports at the time of clinical evaluation. White et al. reassessed the detectability of missed lung cancers in 89 patients⁽³⁰⁾. They included all prior images of missed cancers in their study and were able to include 114 positive images. The CAD system (ONGUARD; Riverain Medical) correctly identified 53 (47%) of the originally missed lung cancers, which indicates the potential of CAD to reduce the number of missed cancers. However, this study only evaluated the stand-alone performance of CAD, which does not include reader response to CAD candidate. It is not clear how many of the tumours detected by CAD would be accepted by readers and at what cost in terms of accepting false positive CAD candidates. Li et al. published the stand-alone performance of the same CAD scheme for a selected group of 34 CT proven T1 cancers that had been missed in the original chest radiography reports⁽³¹⁾. CAD detected 35% of these originally missed lesions. Sensitivity was 45% for the more obvious and 30% for the more subtle cases. The mean number of false positive CAD candidates per image 5.9.

Differentiation of benign from malignant lesion

Differentiating benign from malignant pulmonary nodules on chest radiographs is a difficult and in many cases impossible task. Usually patients with newly identified non-calcified nodules undergo further diagnostic workup that may include CT, PET or biopsy. The superiority of PET-CT for detecting metabolic activity and the ability to perform accurate growth measurements with CT have made early attempts to use morphological features on radiography for nodule differentiation less attractive⁽³²⁾. However, in 2003 Shiraishi et al. presented a CAD algorithm trained to discriminate benign nodules from malignancies with a performance better than even experienced radiologists ($p < 0.002$)⁽³³⁾. In addition to clinical features such as age, sex, and history of the patients, a computerized analysis of morphological features, such as irregularity, density, and contrast had formed the base for this CAD scheme. With CAD the increase in performance was higher for experienced than for inexperienced radiologists, which suggests that CAD and personal skills are complementing effects. Most interestingly, the stand-alone performance of CAD was higher than the performance of each radiologist independent of experience. This demonstrates that radiologists may be reluctant to accept the suggestions of a CAD⁽³³⁾. In a second study the same group of authors tested the combined -detection and classification - task using an advanced CAD scheme and reported similarly significant improvement for both, detection and classification with application of CAD (Az 0.724 vs. 0.778, $p = 0.008$)⁽³⁴⁾. In summary, both detection and classification of T1 tumors has been shown to increase with the application of CAD which is especially appealing for lesions that had been missed during the primary reports.

Limitations of present studies

The vast majority of studies assessed the performance of CAD for a selected study group with a varying prevalence and a varying conspicuity of lesions. The selection of the lesions, with respect to size, location, and overall conspicuity, is tremendously important when interpreting the results. Additional information, such as the fact that the lesions had been originally missed or a relatively low sensitivity below 40%, serve as indicators for a generally low conspicuity of the majority of lesions. Inclusion of many small nodules (5 – 10 mm, based on CT) of which many might be hardly or not at all visible on a chest radiograph limits the assessment of CAD, as well as the inclusion of so obvious lesions that the performance without CAD is already exceeding an Az of 0.9. As recently pointed out in an editorial by D. Gur, the higher the baseline performance is to start with, the more difficult it becomes

to prove a clinically relevant improvement by a certain technique e.g., CAD) ⁽³⁵⁾. The consequence is that the various studies show widely varying levels of improvement, due to the varying level of baseline performances. He also stressed the important point how difficult the appreciation of a 'clinically relevant performance difference' is as one approaches a perfect performance level. While statistics may show significance already for small differences, the perspective in terms of clinical relevance and importance remain frequently unanswered. In studies with a limited and selected group of readers (mostly four to six) the chance that outliers have a great influence on the outcome results is high. Freedman graphically demonstrated the diversity across readers and cases without and with the use of CAD by so called "heat maps" ⁽³⁶⁾. These "heat maps" display three-dimensional data in two dimensions with the third dimension represented by colour. They were found well suited to document the complexity of reader variability as function of lesion conspicuity, experience and level of training, and the interaction between these facts with CAD. This diversity, however, is easily missed in summary statistics, but represents the determinant factor in the individual (clinical) situation. An equally important point is the integration of readers' behaviour. For organizational reasons it is obviously easier to assess the stand-alone performance of a CAD system without including readers in the study. But, it should not be forgotten that these systems are thought to be used as second reader. The highest performance can be expected when readers use CAD as complimentary tool. To be able to use a CAD to its full advantage, readers show a learning curve; they have to become familiar with the potential and limitations of the used CAD algorithm. Readers have to build up a "trust" in order to be able to accept the true positive candidates without a too big loss of specificity by correctly dismissing false positive candidates. More experience has to be gained with respect to the influence of the processing on the performance of CAD. Though there is a tendency towards a consensus to what represents a "good quality" chest radiograph with respect to processing, there remains still room for individually customized version that may drastically differ from the image display in other institutions. Most manufacturers use some type of multi-scale multi-frequency processing which has similar, but not identical effects on the image display. He et al. found a significant impact on the performance of CAD for the detection of nodules when comparing a default processing (parameter 0.0 for structure preference) with a high-pass filtered image (parameter 0.4 for structure preference) and a low-pass filtered image (parameter -0.4 for structure preference) ⁽³⁷⁾. There is so far only one study that aimed to investigate the effect of

CAD in daily practice⁽²⁸⁾. It is thus far the only study which provides some information about the order of magnitude of potential improvement through CAD which can be expected in clinical routine and how many radiographs have to be evaluated for that. The results are valid for a specific group of oncology patients with known extra-pulmonary malignancies. In a clinical setup, establishment of truth remains a problem because not all of these patients can undergo CT and vice versa inclusion of only patients with CT would represent a selection bias. It is therefore not surprising that only for 66% of the original study images, a truth could be established based on follow-up and CT. Prevalence of disease was unusually high with 26% (55/214) of patients with lung nodules. According to the authors, pathology (e.g., nodules) was established in 19 of the 55 patients only with the help of CAD, resulting in an increase of sensitivity from 64% to over 90%. Further clinical evaluation, however, confirmed nodules in only 16 patients, and in only 5 of the 16 patients the nodules were proven to be malignant.

Discussion

Sensitivity versus specificity

Increase in sensitivity and decrease of specificity are somewhat correlated to each other. This is easily understandable if we consider the concept of each reader's "individual threshold" to accept or dismiss a lesion. This concept is true for lesions the reader detected himself as well as for CAD candidates. Even though many, if not most of the false positive CAD candidates, may be easily dismissed, there is a substantial amount of lesions which are difficult to differentiate between projection effects, scarring and a potential malignancy. These lesions increase in number in patients with pre-existing lung disease, such as chronic obstructive lung disease, chronic airways disease or interstitial lung disease. Especially for these patient groups it seems inevitable that a substantial increase in sensitivity associated with CAD goes along with a decrease in specificity. The opposite effect is seen when readers keep the threshold high: a substantial number of correctly indicated lesions by CAD will not be accepted as such. In summary learning effects to get adjusted to the characteristics, e.g., capacities as well as limitations of the CAD algorithm seem to be very important. The detrimental effect of decreasing specificity, or at least substantially decreasing reader confidence, is best studied in lesion-free images ("controls"). Images, previously called not suspicious will now be called suspicious or cases, previously only being characterized as being of low suspicion will now be called "highly suspicious", when

also identified by CAD. The ability of differentiating false from true positive CAD candidates is significantly correlated with the experience of the reader and the amount of anatomic background noise in the source image. Yet, a decrease in specificity with the application of CAD not only applies for inexperienced readers. A quantification of the secondarily induced diagnostic work-up seems warranted with respect to radiation dose, financial aspects, number and invasiveness.

Decrease of intra- and interobserver variability

Freedman et al. demonstrated that most of the lesions newly identified by one radiologist using CAD were actually cases that the radiologist himself or other radiologists would have identified had they interpreted the images without CAD at another point of time⁽³⁶⁾. Even when the application of CAD did not succeed in significantly increasing the sensitivity, in most studies so far it could be shown that it decreases intra- and interobserver variability⁽³⁶⁾.

Value of a high negative predictive value of CAD

The majority of radiographs are indeed done to exclude pathology. It is conceivable that the increase in confidence by CAD to exclude pathology ("I did not see a nodule and CAD also did not find a nodule") will play a substantial role in clinical practice. This depends, of course, on the prevalence of disease in the patient group. However, it has to be noted that the less than 100% sensitivity of CAD requires its use as "second reader".

Options for the future

Computer-aided detection (CAD) has become one of the major research topics in medical imaging and diagnostic radiology. It has been applied to various imaging modalities including CT, MRI, ultrasound, and radiography. Current CAD schemes appear to have the potential to increase the sensitivity for detection of focal lung lesions, such as small tumours and metastases. CAD appears to work best for less experienced readers. At the same time CAD will slightly increase the number of false positive readings. Up to now there is little evidence about how CAD performs in a clinical environment with regard to diagnostic accuracy on one side and negative effects, such as unnecessary follow-up, on the other side. Publications so far tended to focus on the effects of CAD on diagnostic accuracy, its effects on productivity may eventually prove to be equally important. Development and evaluation of these techniques require large databases with validated clinical images. Ideally, common databases should become

available, which can be used to train and test CAD systems, and improve comparison of future commercial systems. This issue is being addressed successfully by several initiatives with the goal to provide a database of web accessible cases for use in CAD research⁽³⁸⁾.

References

- 1 Henschke CI, McMauley DI, Yankelevitz DF, et al. Early lung cancer action project: overall design and findings from baseline screening. *Lancet* 1999; 354:99-105
- 2 Swensen SJ, Jett JR, Hartmann TE, et al. CT screening for lung cancer: five year prospective experience. *Radiology* 2005; 235:259-265
- 3 White CS, Sali AI, Meyer CA. Missed lung cancer on chest radiography and computedtomography:imagingandmedico-legalissues.*JThoracImag*1999;14:63-68
- 4 Muhm JR, Miller WE, Fontana RS, Sanderson DR, Uhlenhoop MA. Lung cancer detected during a screening program using 4-month chest radiographs. *Radiology* 1983; 148:609-615
- 5 Quekel LG, Kessels AG, Goei R, Engelshoven JM. Miss rate of lung cancer on the chest radiograph in clinical practice. *Chest* 1999; 115:720-724
- 6 Shah PK, Austin JH, White CS, et al. Missed non-small cell lung cancers: Radiographic findings of potentially resectable lesions evident only in retrospect. *Radiology* 2003; 226:235-241
- 7 Gavelli G, Giampalma E. Sensitivity and specificity of chest x-ray screening for lung cancer. *Proceedings of the International Conference on Prevention and Early Diagnosis of Lung Cancer*. 1998: 103-108
- 8 Burgess AE, Wagner RF, Jennings RJ. Human signal detection performance for noisy medical images. *IEEE Computer Soc Int Workshop Med Imaging* 1982: 88-105
- 9 Samei E, Flynn MJ, Eyler WR. Detection of subtle lung nodules: relative influence of quantum and anatomic noise on chest radiographs. *Radiology* 1999; 213:727-734
- 10 Kundel HL, Revesz G. Lesion conspicuity, structured noise, and film reader error. *AJR Am J Roentgenol* 1976; 126:1233-1238
- 11 Boynton RM, Bush WR. Recognition of forms against a complex background. *J Opt Soc Am* 1956; 46:758-764
- 12 Revesz G, Kundel HL, Graber MA. The influence of structured noise on the detection of radiologic abnormalities. *Invest Radiol* 1974; 9:479-486
- 13 Bick U, Diekmann F. Digital mammography : what do we and what don't we know. *Eur Radiol* 2007 ; 17 :1931-1942
- 14 Astley SM, Gilbert FJ. Computer-aided detection for screening mammography (review). *Clin Radiol* 2004 ; 59:390-399
- 15 Van Ginneken B, te Haar Romeny BM, Viergever MA. Computer-aided diagnosis in chest radiography: a survey. *IEEE Trans Med Imaging* 2001 ; 20:1228-1241
- 16 Matsumoto T, Doi K, Kano A, Nakamura H, Nakanishi T. Evaluation of the potential benefit of computer-aided (CAD) for lung cancer screenings using photofluorography: analysis of an observer study. *Nippon Acta Radiol* 1993 ; 53:1195-1207
- 17 Warren Burhenne LJ, Wood SA, D'Orsi CJ, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology* 2000 ; 215:554-562
- 18 Chen JJ and White CS. Use of CAD to evaluate lung cancer on chest radiography. *J Thorac Imaging* 2008; 23:93-96
- 19 Kobayashi T, Xu XW, MacMahon H, Metz CE, Doi K. Effect of a computer-aided diagnosis scheme on radiologists' performance in detection of lung nodules on radiographs. *Radiology* 1996; 199:843-848
- 20 MacMahon H, Engelmann R, Behlen FM, et al. Computer-aided diagnosis of pulmonary nodules: results of a large scale observer test. *Radiology* 1999; 213:723-726
- 21 Kakeda S, Moriya J, Sato H, et al. Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system. *AJR Am J Roentgenol* 2004; 182:505-510

- 22 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 characteristics analysis of radiologists' detection of pulmonary nodules. *AJR Am J Roentgenol* 2000; 174:7-74
- 23 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* 2005; 5749:272-280
- 24 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
- 25 Schilham A, van Ginneken B, Loog M. A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database. *Med Image Anal* 2006; 10:247-258
- 26 Hardie RC, Rogers SK, Wilson T, Rogers A. Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. *Med Image Anal* 2008; 12:240-258
- 27 Kasai S, Li F, Shiraishi J, Doi K. Usefulness of computer-aided diagnosis schemes for vertebral fractures and lung nodules on chest radiographs. *AJR Am J Roentgenol* 2008; 191:260-265
- 28 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
- 29 Sakai S, Soeda H, Takahashi N, et al. Computer-aided detection on digital chest radiography: validation test on consecutive T1 cases of resectable lung cancer. *J Digit Imaging* 2006; 19:376-382
- 30 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
- 31 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
- 32 Gietema HA, Schaefer-Prokop CM, Mali W, Groenewegen G, Prokop M. Pulmonary nodules: Interscan variability of semiautomated volume measurements with multisection CT- influences of inspirations level, nodule size, and segmentation performance. *Radiology* 2007; 245:388-394
- 33 Shiraishi J, Abe H, Engelmann R, et al. Computer-aided diagnosis to distinguish benign from malignant solitary pulmonary nodules on radiographs: ROC analysis of radiologists' performance – initial experience. *Radiology* 2003; 227:469-474
- 34 Shiraishi J, Hiroyuki A, Li F, et al. Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs ROC analysis of radiologists' performance. *Acad Radiol* 2006; 13: 995-1003
- 35 Gur D. Imaging technology and practice assessment studies: importance of the baseline or reference performance level. *Radiology* 2008; 247:8-11
- 36 Freedman M, Osicka T. Heat maps: an aid for data analysis and understanding of ROC CAD experiments. *Acad Radiol* 2008; 15:249-259
- 37 He Q, He W, Wang K, Ma D. Effect of multislice processing in digital chest radiography on automated detection of lung nodule with a computer assistance system. *J Digit Imaging* 2008; 21:164170
- 38 MacMahon H. Advanced image processing and computer-aided diagnosis: are we there yet? *J Thorac Imaging* 2008; 23:75-76