Computer-aided Diagnosis: How to Move from the Laboratory to the Clinic[1]

Bram van Ginneken, PhD
Cornelia M. Schaefer-Prokop, MD, PhD
Mathias Prokop, MD, PhD

Computer-aided diagnosis (CAD), encompassing computer-aided detection and quantification, is an established and rapidly growing field of research. In daily practice, however, most radiologists do not yet use CAD routinely. This article discusses how to move CAD from the laboratory to the clinic. The authors review the principles of CAD for lesion detection and for quantification and illustrate the state-of-the-art with various examples. The requirements that radiologists have for CAD are discussed: sufficient performance, no increase in reading time, seamless workflow integration, regulatory approval, and cost efficiency. Performance is still the major bottleneck for many CAD systems. Novel ways of using CAD, extending the traditional paradigm of displaying markers for a second look, may be the key to using the technology effectively. The most promising strategy to improve CAD is the creation of publicly available databases for training and validation. This can identify the most fruitful new research directions, and provide a platform to combine multiple approaches for a single task to create superior algorithms.

[1] From the Department of Radiology, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands (B.v.G., M.P., C.M.S.P.); Image Sciences Institute (B.v.G.) and Department of Radiology (M.P.), University Medical Center, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands; Department of Radiology, Meander Medical Center, Amersfoort, the Netherlands (C.M.S.P.); and Department of Radiology, Academic Medical Center, Amsterdam, the Netherlands (C.M.S.P.). Received September 4, 2009; revision requested October 22; revision received April 12, 2010; accepted April 20; final version accepted May 19; final review by B.v.G. July 25, 2011. **Address correspondence to** B.v.G. (e-mail: *b.vanginneken@rad.umcn.nl*).

A crisis is looming in radiology. The number of imaging studies grows steadily, as well as—may be even more drastically—the number of images per study (1). Radiology is now threatened by its own success: The workload of radiologists has increased dramatically, the number of radiologists still is limited, and health care costs related to imaging are rising fast. New approaches to handle the data explosion are therefore needed. Computer-aided diagnosis (CAD) may hold the key to the problem: If successful, it can speed up the diagnostic process, reduce diagnostic errors, and improve quantitative evaluation.

In this article we consider the definition of CAD broadly, as in "the use of computer algorithms to aid the image interpretation process" (2). Most CAD systems are about detection, which is why CAD can also stand for *computer-aided detection*. CAD is now widely used as a general term, including computerized extraction of quantitative measurements from medical images. From the point of view of algorithm development this is natural, because detection and quantification use similar underlying techniques, and because they are both part of the diagnostic process.

Essentials

- Computer-aided diagnosis (CAD) holds great potential for radiology; its utilization will depend on its ability to speed up the diagnostic process and reduce errors.
- Cost efficiency and reimbursement are important considerations as long as CAD is not yet an integral part of every clinical workstation.
- Many current systems do not yet perform well enough to be considered useful by most radiologists; creating vast databases for training and validation of CAD are the most promising strategies to rapidly improve CAD.
- The standard paradigm of CAD as the image equivalent of a "spellchecker" may not be the optimal use of the technology; using CAD for interpretation, by providing the radiologist with the degree of suspicion for lesions as determined by the computer, may be more effective than only placing markers.
- Radiologists have a key role in CAD development: They need to identify promising application areas, help create high-quality annotated databases for training and validation of CAD systems, and demand that manufacturers embrace open standards so that the best CAD software can be readily installed on any workstation.

CAD and Computing Power

To a large degree, progress in CAD is fueled directly by Moore's law, the doubling of computational power every 2 years (3). This trend has continued for half a century and is not expected to stop for at least another decade. According to futurists like Kurzweil (4), a $1000 computer will have the computational power of a human brain around 2030, and the power of hundreds of human brains 10 years later. Faster and cheaper hardware allows for more intensive computations and for larger training databases. This will generally lead directly to better CAD performance.

However, improvements in software are also needed. First results on the classification of pulmonary lesions on chest radiographs were published in *Radiology* by Lodwick et al in 1963 (5) and he used the term computer-aided diagnosis for the first time in the scientific literature in 1966 (6). Current computers are about 30 million times faster than the ones used by Lodwick, but developing better software algorithms for the detection and classification of pulmonary nodules on chest radiographs is still an active research area (7). It becomes increasingly difficult to improve CAD algorithms, and ever faster computers offer developers more possibilities to explore. The yearly number of publications related to CAD has increased fivefold in this decade compared with the previous one (Fig 1). CAD is on the verge of a breakthrough and in some application areas it can already rival radiologists' performance in terms of sensitivity and specificity. However, in many other areas CAD systems generate many more false-positive findings than their human counterparts, when set to perform at sensitivity levels comparable to those of an experienced radiologist.

Requirements for CAD

CAD has to meet several demands to be used widely in clinical practice. We distinguish four major requirements: *(a)* CAD should improve radiologists' performance. *(b)* CAD should save time. *(c)* CAD must be seamlessly integrated into the workflow. *(d)* CAD should not impose liability concerns and the incremental costs should be negligible or reimbursed.

Most CAD systems today do not meet all requirements, and this is why most applications described in the rapidly growing body of scientific literature on CAD are not widely used in clinical practice.

CAD for Lesion Detection

Current systems for computer-aided detection have been introduced as complementary tools that draw the radiologists' attention to certain image areas that need further evaluation. They do not detect *all* potential lesions, which would allow the radiologist only to focus on the areas identified by the CAD

Published online
10.1148/radiol.11091710 **Content code:** IN

Radiology 2011; 261:719–732

Abbreviations:
CAD = computer-aided diagnosis
FDA = Food and Drug Administration
PMA = premarket approval

Authors stated no financial relationship to disclose.

system. In other words, it is necessary for the radiologist still to evaluate the whole image. However, the systems could detect additional lesions that might have escaped the radiologist's attention.

How Does It Work?

Programming digital computers to understand images is very difficult: One tries to give a step-by-step recipe to tell if an area on an image looks suspicious. To do this, a CAD system breaks the problem into various components. For radiologists who use CAD, it is important to have a basic understanding of these components to be able to understand why the output of a CAD system is sometimes incorrect even if the error is obvious for a human. Unfortunately, most vendors reveal little about the inner workings of their algorithms and supply only a "black box."

Preprocessing.—Many CAD systems start by preprocessing the image data. Scanned images need to be calibrated, data may have to be resampled to a fixed resolution, and noise removal can be applied. The goal of preprocessing is to remove differences between data from different sources or obtained with different protocols. Computers, being blind number crunchers, are easily misled by differences that humans can readily ignore. If a CAD system is trained with and tested only on data from one institution, as is commonly the case for studies reported in the scientific literature, preprocessing may not be necessary, but the results may not be generalized and may not hold true in different settings.

Segmentation.—The second step is segmentation, the division of an image into anatomic regions. Segmentation can be very challenging and is considered the pinnacle of computer vision (8) and one of the most studied areas in medical image analysis (64). Humans rely on a variety of cues, on prior knowledge, and on recognition of related structures to parse an image and immediately see "what is where." Compared with this, current computer algorithms for segmentation are still crude. Figure 2 shows an example of locating the unobscured lung fields on chest radiographs. This case is simple for radiologists because they understand the contents of the image, and immediately notice an unusual amount of air in the colon. Several state-of-the-art computer segmentation methods failed with this case; no methods exist today that are so versatile that they recognize something (air in the colon) that is not part of the task (outlining the lung fields). Computer-aided detection systems usually do not show what they have segmented, arguably because this is not the information the user is primarily interested in. However, incomplete segmentation can make CAD systems miss lesions in the unsegmented areas.

Candidate detection.—Next, a number of locations are identified that merit further attention. These locations are usually referred to as candidates: potential tumors, microcalcifications, polyps, but also regions that may hold diffuse abnormalities. Candidate detection is the most application-specific step in a CAD system. It could consist, for example, of a nodule-enhancing filter followed by thresholding. At this stage, CAD should be very sensitive, while specificity can be low; any lesion that is missed at this stage will not be detected, while improving specificity is the goal of the next stages where each candidate lesion is scrutinized in more detail.

Feature extraction.—Next, each candidate is further analyzed. Almost all systems use the vector space paradigm. This means that for each candidate lesion, a fixed number of characteristics, called features, are computed. Features may be the mean value across the candidate, the standard deviation of the values, the border gradient, or other more complex mathematical descriptors of the candidate and its surroundings. This is a crucial step in a CAD system because now each candidate is represented by a vector, a row of numbers, one for each feature. The feature vector can be represented geometrically by a point in feature space. This feature space has a dimension that is identical to the number of features.

Figure 1

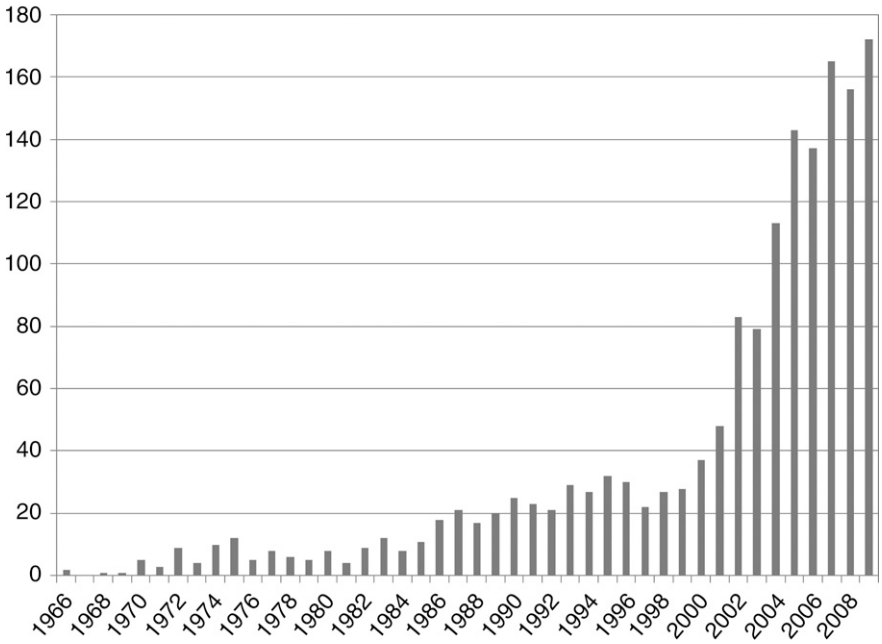

Figure 1: Graph shows number of publications on or related to CAD per year from 1966 through 2009 (last bar). Data were obtained from a PubMed search with search term "computer-aided diagnosis"[Title/Abstract] OR "computer-aided detection"[Title/Abstract] OR "computer-assisted diagnosis"[Title/Abstract] OR "computer-assisted detection"[Title/Abstract].

Figure 2

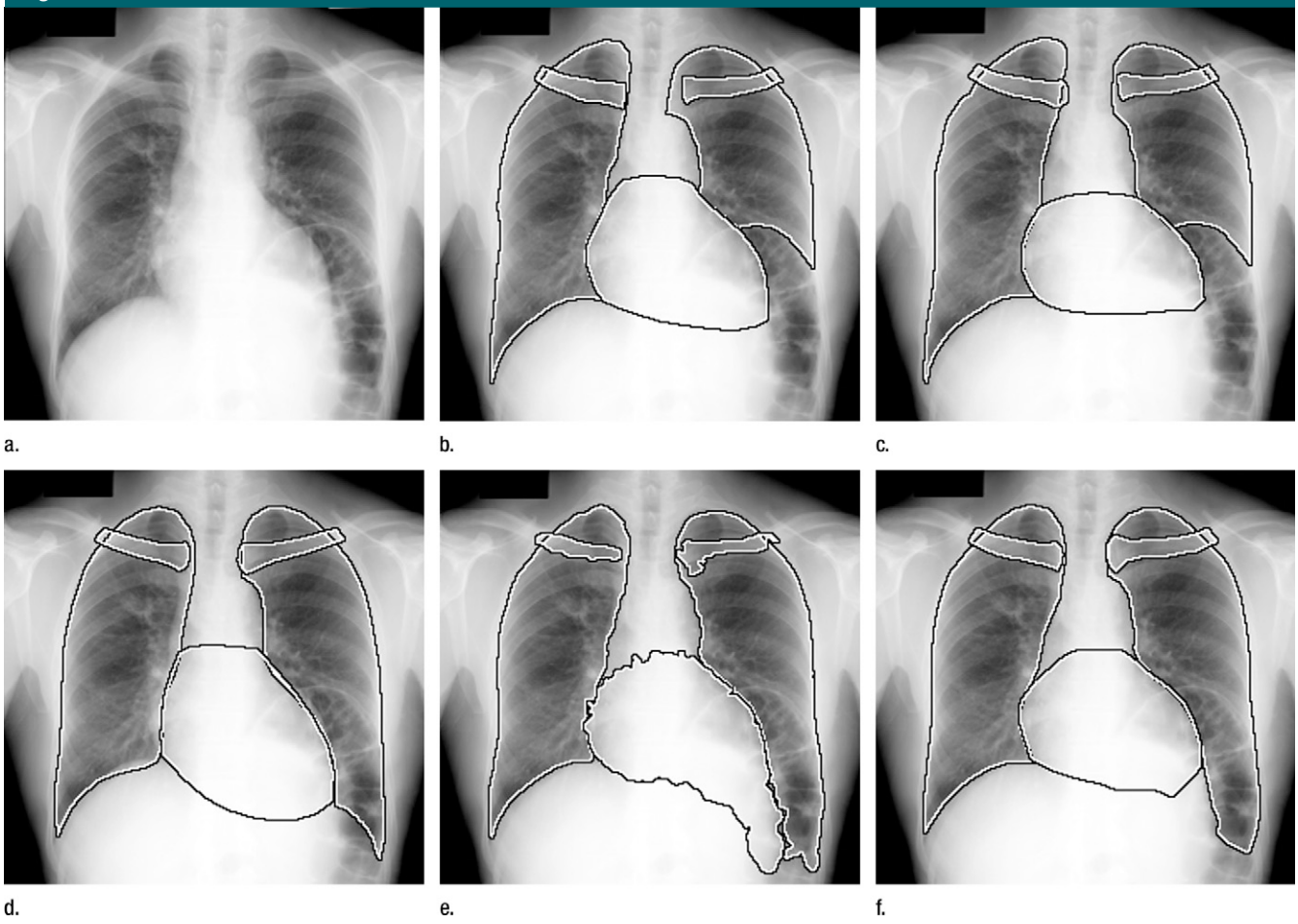

Figure 2: **(a–c)** Chest radiograph and two segmentations of lung fields, heart, and clavicles performed by human operators. The protocol indicated that the contour of the lung field should follow the diaphragm. **(d–f)** Three segmentations produced by state-of-the-art segmentation algorithms: from left to right, active shape models, pixel classification, and active appearance models, respectively (57). In this case the left diaphragm is elevated and there is a large air-distended colon loop directly underneath it. All automatic segmentation methods fail to cope with this unusual situation.

This places the problem within the well-established domain of pattern classification, also known as machine learning. Pattern recognition techniques are widely applied in countless applications outside medical image analysis.

Classification.—In pattern recognition, the problem of identifying regions in feature space where normal and abnormal candidates are located is solved by training a classifier. This requires a training set consisting of correctly classified candidates. Such a training set is usually created by having a human provide a reference standard with correct classifications, for example by indicating lesions on chest radiographs with prior knowledge of their location on a computed tomography (CT) scan obtained in the same patient.

The process is complicated by the fact that some true lesions may reside within regions in the feature space mostly occupied by "non-lesions" or vice versa. A wide variety of well-established classification techniques exist (9,10). A wealth of techniques is also available to automatically select the best features from a large set of potentially useful ones, or to combine features to yield more powerful ones. The major research laboratories in CAD have large computer clusters that are permanently running feature selection processes. Contrary to what is suggested in some studies, there is no single best technique for classification; neural networks (11), support vector machines (12), and Bayesian techniques (10) are all mathematical models that may or may not work well depending on the task at hand. CAD researchers should therefore always experiment with several classifiers.

System output.—The CAD analysis ends when every candidate has been assigned a degree of suspicion by the classifier. All candidates with a degree of suspicion above a threshold, often up to a preset maximum of markers per image data set, are displayed to the radiologist using an arrow, a circle, or a

CAD Systems Approved or Cleared by the FDA in the United States

Name/Company	What It Does	Type of Approval	First and Last Date
Imagechecker/R2 Technology, Sunnyvale, Calif; Hologic, Bedford, Mass	Mass and microcalcification detection on mammograms	PMA	6/1998–9/2007
Logicon caries detection/GA Industries, Rancho Palos Verdes, Calif	Detection of caries on intraoral radiographs	PMA	9/1998–1/2007
Rapidscreen, Onguard/Riverain Medical, Miamisburg, Ohio	Nodule detection on chest radiographs	PMA	7/2001–8/2007
SecondLook/Icad, Nashua, NH,	Mass and microcalcification detection on mammograms	PMA	1/2002–10/2008
LungCare Nodule Enhanced Viewing/Siemens, Erlangen, Germany	Nodule detection and volumetry at chest CT	510(k)	11/2003
MedicLung/MedicSight, London, England	Nodule segmentation and viewing at chest CT	510(k)	12/2003
CT Colonography/General Electric, Fairfield, Conn	Detection of polyps at CT	510(k)	5/2004
Imagechecker-CT/R2 Technology, Sunnyvale, Calif	Detection of pulmonary embolism at chest CT	510(k)	6/2004
Lung CAR/MedicSight, London, England	Nodule detection and volumetry at chest CT	510(k)	7/2004
Colon Car/MedicSight, London, England	Detection of polyps at CT	510(k)	10/2004
Syngo Colonography/Siemens, Erlangen, Germany	Detection of polyps at CT	510(k)	10/2004
IQQA/EDDA, Princeton, NJ	Nodule detection on chest radiographs	510(k)	10/2004
Kodak Mammography CAD Engine/Carestream, Rochester, NY	Mass and microcalcification detection on mammograms	PMA	11/2004–3/2007
Advanced Lung Analysis 2/General Electric, Fairfield, Conn	Nodule detection and volumetry at chest CT	510(k)	11/2004
Syngo Lung CAD/Siemens, Erlangen, Germany	Nodule detection and volumetry at chest CT	510(k)	10/2006
ImageChecker CT CAD/Hologic, Bedford, Mass	Nodule detection and volumetry at chest CT	510(k)	12/2007

Note.—PMA = premarket approval; this type of approval is needed for devices that pose a serious level of risk to the user, and PMA indicates the FDA believes that a new or modified device is safe and effective. 510(k) clearance means that the FDA considers a device "substantially equivalent" to a predicate device. As of 2003, the FDA allowed 510(k) clearance for workstations with integrated CAD capabilities. It is not legal to say that devices with 510(k) clearance have been approved by the FDA. Information in this table was obtained from the FDA Web site using PMA product code MYN and 510(k) product codes OMJ, NEW, and OEB.

color overlay. For commercial systems, this internal threshold can usually not be adjusted by the radiologist and he or she is not provided with the degree of suspicion computed by the CAD system. The CAD output is therefore just a certain number of markers shown on the image that have to be inspected by the radiologist to determine whether the indicated region is considered true or false positive.

Clinical Applications

A wide variety of lesion detection systems exist and have been described in surveys (2,7,13–20) but there is no authoritative overview of commercialized systems available today. The Table lists CAD systems that have received approval or clearance from the U.S. Food and Drug Administration (FDA).

By far the most widely used lesion detection systems are those aimed at mammography breast cancer screening. From sales figures released by manufacturers it can be inferred that around 10000 CAD systems are now in use in the United States. These systems have been steadily improved over the past decade (21) and have been evaluated in large prospective studies (22–24). Mammography CAD systems detect both mass lesions and clusters of microcalcifications. Especially for the latter, CAD systems show high performance and this feature is appreciated by radiologists in screening practice (80). Most computer systems analyze only a single mammogram at a time. Combining information from multiple views and previous examinations is a promising direction for future research, as well as effectively integrating information from multiple modalities that are becoming available for breast cancer detection, such as ultrasonography (US), magnetic resonance (MR) imaging, and digital breast tomosynthesis (14).

Multiple commercial systems are also available for the detection of lung nodules on CT scans or chest radiographs. They have been tested by independent researchers but no large-scale prospective clinical studies have been reported, and these CAD systems are not yet in widespread clinical use. For nodule detection on chest radiographs, multiple studies have shown that the detection performance can be improved using CAD, especially for less-experienced readers, but with variable amounts of decrease in specificity (7). Unlike with CT, with radiography it seems to be much more challenging for the radiologist to effectively differentiate true- from false-positive candidate lesions especially when findings are subtle. Future work should focus on further reduction of the amount of false-positive markers produced by CAD, and the integration of prior radiographs in the CAD analysis. Lung

nodule detection at CT has been very actively developed in the last 10 years (18,81). Results of observer performance studies indicate that accuracy in the detection of lung nodules on chest CT scans can be improved significantly with the use of CAD (14,81). The main challenges for CAD are the detection of nodules with complex vascular attachments and the detection of part-solid and nonsolid nodules (81). Developing CAD schemes to estimate the probability that a pulmonary nodule is malignant is currently an active research field, now that large ongoing screening trials are generating substantial numbers of benign and malignant nodules for training such systems (19).

A number of companies and research groups have developed CAD for the detection of polyps at CT colonography (82). Reported sensitivities of CAD stand alone range from 80% to 100% with two to 15 false-positive markers per scan. Although these results sound very promising, one has to be aware that most of the studies used a small data set for evaluation (82). Not many studies have investigated the effect of CAD on physicians' interpretation of CT colonography studies. One study found that CAD significantly increased per-patient and per-polyp detection and significantly reduced interpretation times (83). Future challenges for CT colonography CAD include achieving robustness for variation in CT scanning techniques and colon preparation methods and combining information from prone and supine scans.

Many other CAD systems are being developed by industry and in academic centers. The Aunt Minnie Web site provides an overview (25) that lists eight systems for mammography, four systems each for breast MR imaging, chest radiography, and CT nodule detection, three systems for CT colonography, two systems each for breast US and positron emission tomography (PET)/CT and single photon emission computed tomography (SPECT)/CT tumor analysis, and one system each for prostate MR imaging analysis, liver CT tumor analysis, and the detection of pulmonary embolism with chest CT. Dozens of other systems have been described in the literature.

CAD Performance for Lesion Detection

The usefulness of CAD depends on the number of true-positive and false-positive markers. Researchers have pointed out that CAD can be useful even if its sensitivity and false-positive rate are below that of the radiologist using the system: The theory is that true-positive CAD markers pointing to lesions initially missed by the radiologist will be accepted by him, while he will not be moved to accept an excessive number of false-positive markers. This theory may not always hold in practice, and this may be one reason why some studies find a significant improvement when readers use CAD, whereas other studies do not find such an effect or find only an increase in sensitivity with a corresponding loss of specificity (7,24). For some applications, for example nodule detection in thoracic CT, researchers have implied that the large number of false-positive markers is not a major problem because these markers can easily be dismissed by the radiologists (26). Even if this statement were true, checking many false-positive markers is likely to increase reading time. Our own clinical experience suggests the following rule of thumb: To be useful, stand-alone performance of CAD should be close to that of an expert radiologist—that is, at the specificity level of an expert, CAD should reach expert sensitivity.

Challenges

Evaluation.—FDA approval was considered mandatory for lesion detection systems from the early days of CAD, but the FDA has not defined a clear policy regarding evaluation requirements for CAD. Since 2004, no more new systems have received PMA approval. Instead, several devices received the much less stringent 510(k) clearance. Manufacturers have described lesion detection algorithms as an advanced visualization filter and received clearance for diagnostic workstations that included these visualization tools. These workstations were considered equivalent by the FDA to other workstations that lacked this functionality. In the past few years hardly any CAD system has received approval, to the general dismay of manufacturers. Last year the FDA organized a hearing to discuss how CAD systems should be evaluated (27).

It is surprisingly difficult to evaluate CAD. Stand-alone performance of CAD without human interaction is a very good indicator of the magnitude of the potential effects of CAD—it determines how many and what kind of lesions can be detected at how many false-positive markers per examination. However, when CAD takes the role of a second reader, the interaction between CAD and the reader will eventually determine the diagnostic impact of the CAD application. It is not possible to quantify this from reader studies alone, because there are many aspects involved that only become apparent once systems are used in clinical routine, such as economic, psychologic, medicolegal consequences, and the health care implications of secondary induced examinations.

As recently pointed out in an editorial by Gur (28), it becomes more difficult to prove a clinically relevant improvement by a certain technique (eg, CAD) the higher the baseline performance of the radiologist is to start with. Studies that evaluate CAD, even for the same application, show widely varying levels of baseline performance (7,28) While statistics may show significance for small differences, the perspective in terms of clinical relevance and importance remains 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.

Interpreting CAD markers.—Simply providing the radiologist with candidate lesions may not represent the most effective way to achieve the goal of CAD—increase sensitivity without a corresponding loss of specificity. Our clinical experience so far shows that it is harder for the radiologist to differentiate between true- and false-positive

markers than one would expect (78). Dependent on the individual acceptance or skepticism level of the reader, he will dismiss true-positive CAD candidate lesions or accept false-positive markers. The trade-off seems to be dependent on the level of the reader's radiologic training but also on his or her experience with the CAD system. Additional processing options, for example rib subtraction and grayscale reversal in chest radiography, might represent the additional "drop of information" that the radiologist needs to successfully differentiate between true- and false-positive lesions. Alternatively, the outcome and the quality of the computer analysis could be conveyed in other ways, for example by showing for each mark reference findings that CAD found to be similar (29).

In Figure 3, taken from Karssemeijer et al (30), the stand-alone performance of a leading commercial mammography mass detection CAD system (ImageChecker v8.0; R2 Technology, Sunnyvale, Calif) that has taken many man-years of development time is plotted. The solid line shows CAD sensitivity as a function of the number of false-positive markers per case. In the same plot the performance of 10 experienced screening radiologists is also plotted (the individual symbols) for various sensitivity levels (the radiologists were asked to mark any finding that they considered even mildly suspicious and rate the degree of suspicion on a scale from 0 to 100). The radiologists can find around 50% of the masses, all of which were prior cancer cases, at false-positive levels ranging from 0.7% for the best radiologists to around 5% for most others. For CAD to find half of the cancers, 15% of the cases contain a false-positive mark, and this percentage may even have been higher for older-generation CAD systems. To achieve high sensitivity, mammography CAD systems are therefore typically set to produce around 0.4 false-positive markers per image (30). At this setting they produce many more false-positive markers than a radiologist who operates in a screening environment. The radiologist has to pick out the few true-positive markers among all these false-positives markers. This is both difficult and time-consuming and could be the reason why many researchers believe that CAD for mass detection on mammograms is not yet good enough to be useful (32).

Figure 3

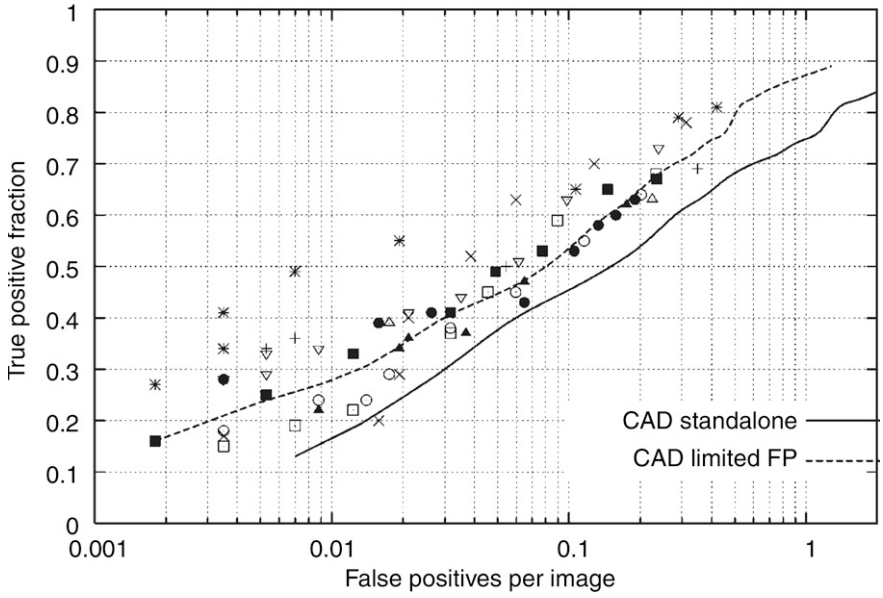

Figure 3: Free-response receiver operating characteristic analysis for mass detection on mammograms. The solid line plots the sensitivity (true-positive fraction) of a commercial CAD program (ImageChecker v8.0; R2 Technology) as a function of false-positive markers per image (horizontal axis, logarithmic scale). The database included 116 prior mammograms of cancer cases and 250 normal cases. Ten expert radiologists were asked to mark and rate on a scale from 0% to 100% all regions that attracted their attention, including those that they would normally not rate suspicious enough for recall. The markers, distinct per radiologist, are the performance levels of the radiologists at different cut-off levels for the provided ratings. Dashed line = stand-alone performance of CAD, taking into account only those regions marked by any of the radiologists. (Reprinted, with permission, from reference 30.)

A New Paradigm for CAD

Conventionally, markers do not indicate the likelihood of suspicion determined by the CAD system. It has been shown that providing this information may be helpful (33). Additionally, one could directly reduce the number of false-positive marks by only showing CAD output when a radiologist clicks on a suspicious region in an image. This simple step can have a dramatic effect, as can be inferred from the dashed line in Figure 3. This is the hypothetical CAD stand-alone performance if all marks on regions that were not marked by any of the 10 expert radiologists, who had to mark even mildly suspicious regions, are removed. If these markers from CAD are disregarded, the CAD system can rival radiologists at this highly challenging task. It was shown in the study by Karssemeijer et al (30) that a simple numerical averaging of the degree of suspicion for a lesion as estimated by a radiologist and by CAD (with the degree of suspicion set to zero when CAD had no detection) resulted in performance that was significantly better than single reading and not significantly different from simulated double reading. Note that this represents a true paradigm shift: CAD is now used as an aid for the *interpretation* of findings, and not as a tool to avoid perceptual oversight. So far, this paradigm has been investigated mainly by one research group and has not yet been

Figure 4

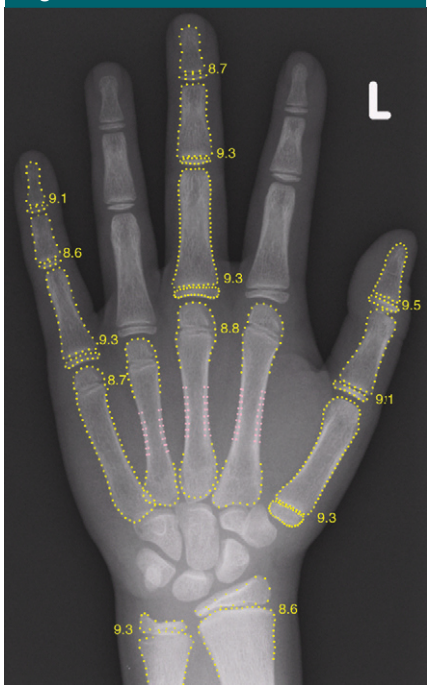

Figure 4: Radiograph of the left hand and wrist analyzed with BoneXpert (42). The program reconstructs the borders of 15 bones and estimates bone age for 13 bones, displayed. These are combined with a nonlinear function to obtain the Greulich Pyle bone age, 9.03 years for this case. Running time for the analysis was 4 seconds.

subject to extensive validation and to other diagnostic tasks than analysis of mammographs.

CAD for Quantification

The amount of quantitative information available on medical images is enormous (34). Computerized quantification may hold more potential than computerized detection. When radiologists are asked to name aspects of their work that are common, time consuming, and could be automated, they usually not do not mention detection but rather documentation and quantification.

How Does It Work?

Quantification and detection are often considered different areas of computerized medical image analysis. This perception is changing. Although many different types of quantitative CAD applications exist, the processing pipeline of these systems is often similar to that of CAD aimed at detection. The crucial difference is the nature of the output: Instead of locations of possible lesions, continuous numbers are produced. Some quantification systems are in fact equivalent to detection systems. For example, automatic calcium scoring to estimate cardiovascular risk (35) requires automatic detection of arterial calcifications. Once detected, the determination of a mass, volume, or Agatston score (36) is a trivial computation.

Data preprocessing is important, especially if differences between scanning protocols need to be corrected. Segmentation is usually a crucially important step. Feature extraction and classification are absent in some applications, especially in older quantitative CAD systems. But "under the hood" of many systems, techniques from pattern recognition are becoming increasingly popular. For example, emphysema quantification from chest CT is traditionally performed with simple thresholding (37), but texture analysis that includes training data, feature extraction, and classification is an alternative way of quantification (38,39) that is less prone to sources of errors and may result in measurements with more prognostic value.

Clinical Applications

There are many systems available for quantitative image analysis (34) and a complete overview is outside the scope of this article. Quantification is applied to cardiac function and vascular abnormalities with MR and CT, brain MR and CT, CT oncology, and other modalities such as US and bone densitometry (25). A large area is quantitative perfusion imaging, with applications in oncology (breast, prostate) and brain (stroke). Another important application is the estimation of size and growth rate of tumors and other lesions. Furthermore, there are systems that quantify the extent of particular diseases, such as chronic obstructive pulmonary disease and interstitial lung disease. They can be used in clinical practice, but are also important in research trials.

There exist many radiologic scoring systems. Reliable automation of established scoring systems has enormous potential to standardize and improve radiology. Many scoring systems are complicated and are known to result in large interobserver variability. Other scoring systems are simpler and therefore a better target for quantitative CAD. We discuss one representative example in more detail: bone age assessment determined from a radiograph of the left hand and wrist. Many attempts to automate bone age assessment have been made, and in 1994 Tanner et al (40) introduced CASAS, an interactive program that they claimed solved the problem of bone age assessment by computer. CASAS was thoroughly evaluated (40,41) and found to reduce interobserver variability. But the program was impractical to use because each bone had to be located manually by the user. Moreover, the automated staging of the bones was based on fairly simple image analysis and performance was not impressive. Thus, CASAS did not meet radiologists' requirements for CAD products and was never widely used.

Automated bone age assessment continued to attract research interest. Recently, Thodberg et al (42) introduced BoneXpert, a program that is fully automatic and processes an image in about 5 seconds with standard PC hardware (Fig 4). It locates the borders of 15 bones on the radiograph and assigns an intrinsic bone age to 13 bones, using a set of 30 features that describe the shape of the bones, the density and the texture that are sensitive to epiphyseal fusion (Fig 4). A nonlinear model recovers the Greulich and Pyle bone age (43) from these intrinsic bone ages. Evaluation studies on 405 and 1097 radiographs found that the system processed over 99% of images automatically without errors (44,45). It seems that Tanner's claim of 1994 is now a reality.

CAD Performance for Quantification

The rule of thumb for lesion CAD also applies to quantification: It should be as good as a radiologist. This rule applies to applications where an external reference standard is not available and

manual quantification is a viable but time-consuming alternative to automated methods.

Performance varies substantially between systems. As an example we consider nodule volumetry. De Hoop et al (46) recently published a direct comparison of six commercially available algorithms for volumetry of solid lung nodules on low-dose thoracic CT scans. They visually inspected the results for each pair of nodules, on a first and second scan. Three commercial systems produced an inadequate segmentation for at least one of the two instances of each nodule in 22%–30% of all cases. The best performing system, however, did this in only 2% of the cases.

Challenges in Quantitative CAD Development

Data Preprocessing and Evaluation

Preprocessing is especially important for quantification. For example, when emphysema is quantified with CT, differences in dose, section thickness, and reconstruction algorithm all affect the emphysema scores substantially (47). In general, to reduce measurement variability, especially in longitudinal studies, applying the same scan and reconstruction parameters is important. In practice, variations will remain and proper preprocessing may be able to reduce these effects (48). A major goal is to develop methods that standardize image quality, in terms of resolution, artifacts, and noise levels.

Data preprocessing can never eliminate all sources of error that play a role in quantitative measurements. Evaluation of quantitative CAD techniques should be focused on measuring these errors. Radiologists must be aware that numbers obtained from the quantitative CAD techniques cannot be trusted blindly. Diagnostic workstations often provide numbers with many digits and therefore seemingly high precision, but in reality the error margin of these numbers is often large, difficult to determine, and almost never listed. It is important to conduct studies that test the repeatability of quantitative measurements, whether manually or automatically made, and this has been done, for example, in emphysema quantification (49), nodule volumetry (46,50), and calcium scoring (51). From such studies the minimum increase in the quantified measure that corresponds to significant change can be deduced, although one must realize that these studies often do not measure the effect of all sources of errors that may be relevant in complex situations such as multicenter trials. This issue makes proper evaluation of the accuracy and reproducibility of quantitative image analysis extraordinarily complex.

Assessing the accuracy and repeatability of quantitative measures is only one aspect of evaluation. Eventually one needs to convince practitioners that using a quantitative measure will improve patient care. In this respect, the proper choice for a metric is important, as pointed out by Boone (34). For example it is clinically more relevant to show that the risk for cardiovascular disease can be predicted more accurately than that the calcified volume in the coronary arteries can be quantified more precisely.

User Interaction

Ideally, quantification should be performed fully automatically. The main reason that most radiologic scoring systems are not used routinely in clinical practice is that they are too time-consuming and cumbersome to apply. In a comparison of four commercially available systems for emphysema quantification (52) it was found that the median time required to process one scan ranged from 16 to 105 minutes and this time was spent mainly in interactively adjusting the lung segmentation needed to make the volumetric measurements. This obviously precludes routine clinical use.

If human verification is required, this should ideally only require a quick glance. Interaction should be intuitive, and results should be available instantaneously. For many tasks reliable fully automatic solutions are not yet available. For the six nodule volumetry systems discussed previously (46), the 98% success rate with the best system was achieved with a simple and effective user interaction in case automatic processing failed. Without any interaction, the two best methods in the study had a failure rate of around 15%.

Workflow Integration

For radiologists to use CAD it must be integrated seamlessly in their workflow: The output of the algorithm should be available within their workstation, preferably at the push of a single button (34,53). The first steps in this direction are being made. Several vendors are marketing CAD servers. These are computers within the network of the radiology department that inspect the Digital Imaging and Communications in Medicine (DICOM) descriptors of all studies, process relevant scans in the background, and store CAD results as DICOM structured reports or, less than ideal, as a copy of the original images with the results of CAD, for example markers or measurement results, superimposed (54,55). In theory, CAD results stored in a standardized format could be displayed by any workstation and a consortium called Integrated Healthcare Enterprise is working on such standards (56). In practice this is not yet the case. Integrating quantitative CAD tools in a diagnostic workstation is more complicated than including CAD detection markers that only require static display. This is especially true if user interaction is deemed necessary. Automatic inclusion of the quantitative information in the radiologic report is essential.

Radiologists, as workstation users, together with patient advocacy groups and regulatory agencies, should insist that vendors allow such interoperability and embrace open standards. Workstation vendors should realize that their product has more value if it allows integration of any CAD product, even a product from a competitor. A useful analogy is diagnostic workstations and mobile phone devices. Some mobile phone manufacturers have created a highly successful market for small applications that enhance a phone's capabilities. This materialized

because they opened their systems to software created by third party companies, instead of bundling only their own software. All parties, phone manufacturers, third-party developers and end users, benefit from this model. If such a market were created for diagnostic workstations, even simple CAD applications that under the current situation would not be profitable to develop could be deployed widely and save radiologists' time. An example is the automatic determination of the cardiothoracic ratio on chest radiographs. Automating such a measurement is feasible (57) but this CAD application is too small to be marketed as a new product. However, as a plug-in or "app" it might be widely used.

Future of CAD Development

Without doubt, CAD technology will be applied to many new tasks. But most existing CAD systems also have ample room for improvement. We believe that evaluation, database collection, and collaborative efforts to design superior CAD systems are the most promising strategies to make rapid progress.

New Opportunities

Despite the increased interest in CAD research, many potential applications have not yet been addressed. In 1991, MacMahon and colleagues (58) investigated how often a range of subtle abnormal findings occur on chest radiographs. Nodules, the only finding for which CAD is commercially available, ranked only fifth on this list. There are many other diagnostic tasks in chest radiography that could benefit from CAD that have not been studied in detail and for which no commercial systems are available. Examples are the detection of catheter tips, subtle infiltrates, pneumothorax, cardiac abnormalities, tuberculosis, pneumonia, and emphysema (13). Summers (53) presented a road map for CAD of chest CT findings in 2003, and although some items on his list have been addressed, such as automatic detection of coronary calcifications (35), many tasks in this road map have not been the topic of any published study to date, including quantification of pleural effusions, detection of subcutaneous nodules, pneumothorax, and bone metastases.

Outside the main areas of CAD (breast, lung, colon) there are many opportunities in skeletal and neuroimaging (15) and in areas outside radiology, such as retinal imaging (31) and dermatologic imaging (63).

Translating Expert Knowledge into Features

The main challenge for algorithm developers is to translate expert knowledge from radiologists (eg, a spiculated nodule is more likely to be malignant) into effective computable features. No matter how complicated the features are, in this translation process most information from the original images is lost. A CAD system is a funnel, at every step the computer reduces its representation of an image to a smaller set of numbers. Limited as this approach may seem, it is the most successful approach available. In the words of Drew McDermott (59), an eminent researcher in artificial intelligence: "What you get for fitting into this straightjacket is that your problem actually gets solved. I expect this trend to continue. Everything worth doing will turn out to be possible with simple representations."

The result is that CAD systems today address single, isolated tasks by particular, painstakingly designed sequences of number crunching. Even if this solves the problem, as McDermott suggests, it has two disadvantages. First, it is time consuming to design a new CAD application because no general theory or recipe on how to do this is available; much of it is more art and engineering than science: For every step in a CAD system, the designer faces multiple choices, and four steps with four options each already gives 256 different possibilities. Second, it does not lead to computers that truly understand images as a whole and can perform high-level reasoning about the content of an image comparable to a radiologist. Consequently, CAD systems often make mistakes that no human would make, as illustrated in Figure 2. True image understanding by machines, following Turing's (60) famous paradigm from 1950 (1. Construct a teachable machine. 2. Subject it to a course of education.), is well beyond the horizon today. When Bill Gates famously said (61) that "if you invent a breakthrough in artificial intelligence, so machines can learn, that is worth 10 Microsofts" he did not add that false promises of breakthroughs have given the term artificial intelligence such a bad name that scientists and engineers have shunned it for decades.

CAD applications will continue to be improved, steadily but slowly. With so many groups entering the field, as is evident from the increasing number of publications, it is essential to know where the best systems fail to speed up the rate of improvement. Open challenges should be clearly identified for each particular task, so that new research is incremental to previous work, instead of merely reproducing it.

Measuring Performance

This leads to the major bottleneck for improving CAD today: It is generally unclear what the best system or approach for a given task is. Even algorithm developers have no idea, let alone potential users. One would expect careful evaluation of the performance of CAD systems to be an active research area, so that radiologists can make informed choices about which CAD system to use and algorithm designers can decide which techniques are most promising for further research. Unfortunately, this is not the case; studies that compare different CAD systems are rare.

Large prospective studies, such as those that have been carried out for mammography CAD (22–24) evaluate only a single CAD system. Suppose that a large multicenter study X shows a modest benefit of using CAD system A in clinical practice. System A is FDA approved and can start being commercialized. While study X is ongoing, another system, B, is completed, and when B is compared with A (which would almost certainly be performed using a different and smaller data set

than the one used in multicenter study X), it achieves significantly higher sensitivity levels at the same false-positive levels. System B would have to undergo a larger multicenter trial as well to truly prove its superiority to A. Since system A has already achieved approval and sells well, the resources to repeat a large multicenter study for system B are unlikely to be granted.

This scenario is not hypothetical; the stand-alone performance of newer, non–FDA-approved versions of CAD systems is frequently substantially higher than that of FDA-approved older systems. Li et al (62) compared FDA-approved version 1.0 of a CAD software for nodule detection on chest radiographs with version 3.0, which was not FDA approved, and reported an increase in sensitivity from 53% to 67% with a simultaneous decrease in the false-positive rate from 5.7 to 2.7 marks per image.

It is not practical to design large-scale evaluation studies where the performance of radiologists with and without CAD is tested for multiple CAD systems. This may not be necessary. Karssemeijer et al (30,65) have shown with a large database of screening mammograms where the annotations of 10 expert radiologists were available that interpretation with CAD can be simulated. They combined the suspicion rating assigned to each finding by the radiologist with the CAD output at the area of that finding. Moreover, the effect of double reading and the effect of combining CAD systems can be assessed. More research is needed to determine if such simulations can accurately predict the outcome of actual observer studies.

Comparing CAD Algorithms

Large databases, representative of clinical practice and carefully annotated, provide the basis for valid comparative studies. The need for publicly available reference databases is widely acknowledged in the CAD research community, and substantial efforts to address it are underway. An excellent example is the work of the Lung Image Database Consortium, five research groups who collect a database of annotated thoracic CT scans (66). The number of cases available has recently been increased to 400 (67) and it has been announced that a total of 1000 cases will be released. The database has been used in several studies for evaluation of CAD algorithms. Unfortunately, different researchers used different selections of cases and methodologies for evaluation. In the future such databases should be randomized in a part that is made available for training, together with all annotations, and a "secret" part set aside for evaluation. Teams could upload their software to a central organization that processes the test scans and reports results of all teams, evaluated according to the same protocols.

Such studies, where multiple algorithms are benchmarked on a single data set, are rapidly gaining popularity in medical image analysis. In 2007, the first contests of this kind were organized (68) focused on segmentation. In total, nine challenges have now been organized and for two of these the results have been published, on the segmentation of the liver on abdominal CT scans (69) and the extraction of the coronary artery tree at CT angiography (70). Recently also CAD contests have been started for lung nodule detection (71) and lung nodule volumetry (72) at thoracic CT, and the detection of microaneurysms on retinal images (73).

Better than Comparing: Combining

The full potential of such challenges can be appreciated by considering the successful Netflix Prize (74,75). In October 2006, DVD rental company Netflix announced it would award $1 million to whomever created a movie-recommending algorithm that outperformed the company's own algorithm by 10%. Netflix made a training database available that was vastly larger than any data set previously accessible to researchers. Within 2 weeks, three results were submitted that were superior to Netflix's own software. During the second and third year of the contest it became increasingly difficult to improve performance and the biggest gains came from combining different approaches to the problem. In July 2009, a coalition of AT&T researchers, a group from Austria, and a team from Montreal sent in a result that qualified for the Netflix Prize. According to the contest rules, the other teams had 30 days to improve on this result. A large group of remaining teams, dubbed The Ensemble, managed to do this by blending their algorithms with a method similar to that used by Karssemeijer et al (30) to combine radiologists' and CAD output. The Netflix Prize has generated an enormous interest in the nascent research field of recommendation systems, with over 40 000 submissions from more than 5000 teams, and has generated a multitude of important research results (84,85). Above all, it has demonstrated that a careful combination of a large number of different algorithms is essential to build a superior system.

We see no reason why a similar approach to the development of a CAD system would not be as successful. Thus, collecting a large annotated database and making it publicly available can provide researchers with training data and allow objective algorithm comparisons. It can also attract more research groups into the field and develop a CAD system that combines a multitude of approaches and will vastly outperform the algorithms available today. The wisdom of crowds (76) also applies to CAD.

Recent results support this hypothesis. In the previously mentioned liver segmentation contest, averaging the output of the five best automatic systems outperformed each individual system and obtained results very close to the best interactive system that participated in the challenge, a virtual reality segmentation system used by liver surgeons (69,77). Figure 5 illustrates how a combination of five systems for nodule detection on thoracic CT scans, submitted to the ongoing ANODE09 contest (71), performs much better than any of five systems individually. Within the pattern recognition and machine learning communities, classifier combination has been studied extensively (79), but in diagnostic image analysis, this notion is relatively new. The best

Figure 5

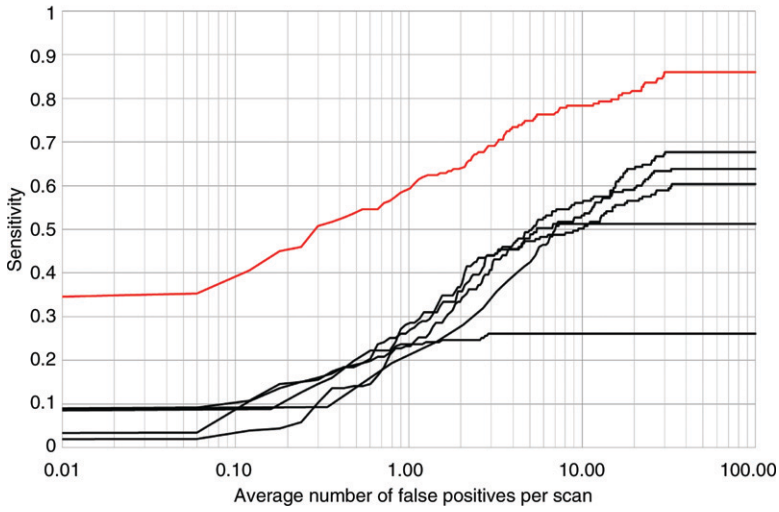

Figure 5: Performance of five different CAD systems for the detection of pulmonary nodules with CT data (black solid lines). Sensitivity (vertical axis), the part of true nodule marked by zCAD, is plotted as a function of the number of false-positive markers per scan (horizontal axis, logarithmic scale). All systems were evaluated in the same way by using 50 CT scans from the ANODE09 data set. Four systems have been developed in academia; one system is a commercially available CAD system. Red line = performance of a combination of the five systems. The combined system is superior to any of the individual systems. (Adapted and reprinted, with permission, from reference 71.)

CAD system for any relevant task is unlikely to be the result of research carried out in a single research laboratory. Instead, research groups should collaborate and together create a system that combines many different approaches in an efficient manner.

Summary and Outlook

It is extraordinarily difficult to develop computer algorithms that analyze medical images with a performance level comparable to that of human experts. The main reason for that is that CAD problems are multifaceted; they require dedicated image preprocessing, segmentation, algorithms to pick up suspect regions, mathematical descriptors of these regions, a classification stage, and, last but not least, a large training database of normal and abnormal cases that is representative of what the algorithm will encounter in the clinic. There is no lack of ambition in the field; most research has focused on tasks that are also very difficult for radiologists. To crack these challenges requires world-class expertise in image processing, machine learning, medical physics, and radiology—CAD is a truly interdisciplinary research field.

The availability of large high-quality databases would tremendously lower the threshold for outsiders to enter the field. Many of the best minds in machine learning are interested in medical applications, but lack access to the data needed to compete with research laboratories and industry that work closely with radiologists. When the results of multiple algorithms on large representative databases are available, it is possible to study the effect of combining multiple algorithms. The power of classifier combination and algorithm blending has been demonstrated in many pattern recognition applications. Results so far indicate that CAD is no exception.

Another major reason why CAD is not yet widely used is that integration in clinical workflow poses many practical challenges. Radiologists must urge manufacturers to open their picture archiving and communication system (PACS) workstations for third-party CAD software developers. They should be able to download and directly use CAD software in their own PACS environment, as easily as they can today download small applications that enhance the capabilities of their mobile phone. This scenario will require the development of new standards.

The number of potential CAD applications is huge. New ways of using CAD technology, extending the paradigm of providing markers for a second look, are an important future research direction. If it is easy to work on CAD development, because an accessible system to validate and combine algorithms has been created, and if high-performing systems that will surface can be easily plugged into all clinical viewing software, then CAD will thrive. Radiologists will become more productive and make less errors because they will have more time to think about the interpretation of their findings. They will benefit, and so will their patients.

References

1. Rubin GD. Data explosion: the challenge of multidetector-row CT. Eur J Radiol 2000; 36(2):74–80.
2. Giger ML, Karssemeijer N, Armato SG 3rd. Computer-aided diagnosis in medical imaging. IEEE Trans Med Imaging 2001;20(12): 1205–1208.
3. Moore's Law. *http://en.wikipedia.org/wiki/Moore's_law*. Accessed August 26, 2009.
4. Kurzweil R. The singularity is near: when humans transcend biology. New York, NY: Viking Penguin, 2005.
5. Lodwick GS, Keats TE, Dorst JP. The coding of Roentgen images for computer analysis as applied to lung cancer. Radiology 1963; 81:185–200.
6. Lodwick GS. Computer-aided diagnosis in radiology. A research plan. Invest Radiol 1966; 1(1):72–80.
7. De Boo DW, Prokop M, Uffmann M, van Ginneken B, Schaefer-Prokop CM. Computer-aided detection (CAD) of lung nodules and small tumours on chest radiographs. Eur J Radiol 2009;72(2):218–225.

8. Jain AK. Fundamentals of digital image processing. Upper Saddle River, NJ: Prentice Hall, 1989.
9. Duda RO, Hart PE, Stork DG. Pattern classification. 2nd ed. New York, NY: Wiley, 2001.
10. Bishop CM. Pattern recognition and machine learning. Secaucus, NJ: Springer, 2007.
11. Duin RP. Superlearning and neural network magic. Pattern Recognit Lett 1994;15(3):215–217.
12. Bennett KP. Support vector machines: hype or hallelujah? SIGKDD Explor Newslett 2000;2(2):1–13.
13. van Ginneken B, Hogeweg L, Prokop M. Computer-aided diagnosis in chest radiography: beyond nodules. Eur J Radiol 2009;72(2):226–230.
14. Giger ML, Chan HP, Boone J. Anniversary paper: History and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. Med Phys 2008;35(12):5799–5820.
15. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph 2007;31(4-5):198–211.
16. Nishikawa RM. Current status and future directions of computer-aided diagnosis in mammography. Comput Med Imaging Graph 2007;31(4-5):224–235.
17. van Ginneken B, ter Haar Romeny BM, Viergever MA. Computer-aided diagnosis in chest radiography: a survey. IEEE Trans Med Imaging 2001;20(12):1228–1241.
18. Sluimer IC, Schilham AM, Prokop M, van Ginneken B. Computer analysis of computed tomography scans of the lung: a survey. IEEE Trans Med Imaging 2006;25(4):385–405.
19. Chan HP, Hadjiiski L, Zhou C, Sahiner B. Computer-aided diagnosis of lung cancer and pulmonary embolism in computed tomography-a review. Acad Radiol 2008;15(5):535–555.
20. Boyer B, Balleyguier C, Granat O, Pharaboz C. CAD in questions/answers: review of the literature. Eur J Radiol 2009;69(1):24–33.
21. Roehrig J. The manufacturer's perspective. Br J Radiol 2005;78(Spec No 1):S41–S45.
22. Dean JC, Ilvento CC. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. AJR Am J Roentgenol 2006;187(1):20–28.
23. Brem RF. Clinical versus research approach to breast cancer detection with CAD: where are we now? AJR Am J Roentgenol 2007;188(1):234–235.
24. Noble M, Bruening W, Uhl S, Schoelles K. Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. Arch Gynecol Obstet 2009;279(6):881–890.
25. Aunt Minnie Buyer's Guide. *http://www.auntminnie.com/index.asp?sec=wdp*. Accessed August 26, 2009.
26. Das M, Mühlenbruch G, Mahnken AH, et al. Small pulmonary nodules: effect of two computer-aided detection systems on radiologist performance. Radiology 2006;241(2):564–571.
27. CDRH Advisory Meeting Materials Archive. *http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfAdvisory/details.cfm?mtg=659*. Accessed August 26, 2009.
28. Gur D. Imaging technology and practice assessment studies: importance of the baseline or reference performance level. Radiology 2008;247(1):8–11.
29. Horsch K, Giger ML, Vyborny CJ, Lan L, Mendelson EB, Hendrick RE. Classification of breast lesions with multimodality computer-aided diagnosis: observer study results on an independent clinical data set. Radiology 2006;240(2):357–368.
30. Karssemeijer N, Otten JD, Rijken H, Holland R. Computer aided detection of masses in mammograms as decision support. Br J Radiol 2006;79(Spec No 2):S123–S126.
31. Abràmoff MD, Niemeijer M, Suttorp-Schulten MS, Viergever MA, Russell SR, van Ginneken B. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. Diabetes Care 2008;31(2):193–198.
32. Nishikawa RM. Computer-aided detection, in its present form, is not an effective aid for screening mammography. For the proposition. Med Phys 2006;33(4):811–812.
33. Jiang Y, Nishikawa RM, Schmidt RA, Toledano AY, Doi K. Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications. Radiology 2001;220(3):787–794.
34. Boone JM. Radiological interpretation 2020: toward quantitative image assessment. Med Phys 2007;34(11):4173–4179.
35. Isgum I, Rutten A, Prokop M, van Ginneken B. Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease. Med Phys 2007;34(4):1450–1461.
36. Rumberger JA. Coronary artery calcium scanning using computed tomography: clinical recommendations for cardiac risk assessment and treatment. Semin Ultrasound CT MR 2008;29(3):223–229.
37. Müller NL, Staples CA, Miller RR, Abboud RT. "Density mask". An objective method to quantitate emphysema using computed tomography. Chest 1988;94(4):782–787.
38. Xu Y, Sonka M, McLennan G, Guo J, Hoffman EA. MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. IEEE Trans Med Imaging 2006;25(4):464–475.
39. Sluimer IC, Prokop M, Hartmann I, van Ginneken B. Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution CT of the lung. Med Phys 2006;33(7):2610–2620.
40. Tanner JM, Oshman D, Lindgren G, Grunbaum JA, Elsouki R, Labarthe D. Reliability and validity of computer-assisted estimates of Tanner-Whitehouse skeletal maturity (CASAS): comparison with the manual method. Horm Res 1994;42(6):288–294.
41. Albanese A, Hall C, Stanhope R. The use of a computerized method of bone age assessment in clinical practice. Horm Res 1995;44(Suppl 3):2–7.
42. Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. IEEE Trans Med Imaging 2009;28(1):52–66.
43. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. 2nd ed. Stanford, Calif: Stanford University Press, 1959.
44. van Rijn RR, Lequin MH, Thodberg HH. Automatic determination of Greulich and Pyle bone age in healthy Dutch children. Pediatr Radiol 2009;39(6):591–597.
45. Martin DD, Deusch D, Schweizer R, Binder G, Thodberg HH, Ranke MB. Clinical application of automated Greulich-Pyle bone age determination in children with short stature. Pediatr Radiol 2009;39(6):598–607.
46. de Hoop B, Gietema H, van Ginneken B, Zanen P, Groenewegen G, Prokop M. A comparison of six software packages for evaluation of solid lung nodules using semi-automated volumetry: what is the minimum increase in size to detect growth in repeated CT examinations. Eur Radiol 2009;19(4):800–808.
47. Boedeker KL, McNitt-Gray MF, Rogers SR, et al. Emphysema: effect of reconstruction algorithm on CT imaging measures. Radiology 2004;232(1):295–301.

48. Schilham AM, van Ginneken B, Gietema H, Prokop M. Local noise weighted filtering for emphysema scoring of low-dose CT images. IEEE Trans Med Imaging 2006;25(4):451–463.

49. Gietema HA, Schilham AM, van Ginneken B, van Klaveren RJ, Lammers JW, Prokop M. Monitoring of smoking-induced emphysema with CT in a lung cancer screening setting: detection of real increase in extent of emphysema. Radiology 2007;244(3):890–897.

50. Wormanns D, Kohl G, Klotz E, et al. Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol 2004;14(1):86–92.

51. Detrano RC, Anderson M, Nelson J, et al. Coronary calcium measurements: effect of CT scanner type and calcium measure on rescan reproducibility—MESA study. Radiology 2005;236(2):477–484.

52. Heussel CP, Achenbach T, Buschsieweke C, et al. Quantification of pulmonary emphysema in multislice-CT using different software tools [in German]. Rofo 2006;178(10):987–998.

53. Summers RM. Road maps for advancement of radiologic computer-aided detection in the 21st century. Radiology 2003;229(1):11–13.

54. Noumeir R. Benefits of the DICOM structured report. J Digit Imaging 2006;19(4):295–306.

55. Zhou Z, Liu BJ, Le AH. CAD-PACS integration tool kit based on DICOM secondary capture, structured report and IHE workflow profiles. Comput Med Imaging Graph 2007;31(4-5):346–352.

56. Integrating the Healthcare Enterprise. *http://www.ihe.net/*. Accessed September 6, 2009.

57. van Ginneken B, Stegmann MB, Loog M. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Med Image Anal 2006;10(1):19–40.

58. MacMahon H, Montner SM, Doi K, Liu KJ. The nature and subtlety of abnormal findings in chest radiographs. Med Phys 1991; 18(2):206–210.

59. Simon HA, Bibel W, Bundy A. AI's greatest trends and controversies. IEEE Intell Syst 2000;15(1):8–17.

60. Turing AM. Can a machine think? Mind 1950;59(236):433–460.

61. Lohr S. Microsoft, amid dwindling interest, talks up computing as a career. New York Times, March 1, 2004.

62. Li F, Engelmann R, Doi K, Macmahon H. True detection versus "accidental" detection of small lung cancer by a computer-aided detection (CAD) program on chest radiographs. J Digit Imaging 2010;23(1):66–72.

63. Maglogiannis I, Doukas CN. Overview of advanced computer vision systems for skin lesions characterization. IEEE Trans Inf Technol Biomed 2009;13(5):721–733.

64. Sonka M, Hlavác˘ V, Boyle R. Image processing, analysis, and machine vision. 3rd ed. Toronto, Canada: Thomson Learning, 2007.

65. Karssemeijer N, Otten JD, Verbeek AL, et al. Computer-aided detection versus independent double reading of masses on mammograms. Radiology 2003;227(1):192–200.

66. Armato SG 3rd, McLennan G, McNitt-Gray MF, et al. Lung image database consortium: developing a resource for the medical imaging research community. Radiology 2004;232(3):739–748.

67. National Biomedical Imaging Archive. *http://ncia.nci.nih.gov/*. Accessed August 26, 2009.

68. van Ginneken B, Heimann T, Styner M. 3D segmentation in the clinic: a grand challenge. In: van Ginneken B, Heimann T, Styner M, eds. 3D segmentation in the clinic: a grand challenge. Brisbane, Australia: MICCAI, 2007; 7–15.

69. Heimann T, van Ginneken B, Styner MA, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans Med Imaging 2009;28(8):1251–1265.

70. Schaap M, Metz CT, van Walsum T, et al. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. Med Image Anal 2009;13(5):701–714.

71. The ANODE09 Competition for Nodule Detection in Chest CT. *http://anode09.isi.uu.nl*. Accessed August 26, 2009.

72. Volcano'09 Challenge. *http://www.via.cornell.edu/challenge/*. Accessed August 26, 2009.

73. Retinopathy online challenge. *http://roc.healthcare.uiowa.edu/*. Accessed August 26, 2009.

74. Netflix Prize. *http://www.netflixprize.com/*. Accessed August 26, 2009.

75. Ellenberg J This psychologist might outsmart the math brains competing for the Netflix prize. Wired, February 2008.

76. Surowiecki JM. The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. New York, NY: Little, Brown, 2004.

77. Reitinger B, Bornik A, Beichel R, Schmalstieg D. Liver surgery planning using virtual reality. IEEE Comput Graph Appl 2006;26(6):36–47.

78. De Hoop B, De Boo D, Gietema HA, Van Hoorn F, Mearadji B, Schaefer-Prokop CM. CAD for the detection of tumors in CXRs of smokers: results of an observer L-ROC study [abstr]. In: Radiological Society of North America Scientific Assembly and Annual Meeting Program. Oak Brook, Ill: Radiological Society of North America, 2009; 533–534.

79. Kittler J, Hatef M, Duin RP, Matas J. On combining classifiers. IEEE Trans Pattern Anal Mach Intell 1998;20(3):226–239.

80. Karssemeijer N, Bluekens AM, Beijerinck D, et al. Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. Radiology 2009;253(2):353–358.

81. Lee N, Lain AF, Marquez G, Levsky JM, Gohagan JK. Potential of computer-aided diagnosis to improve CT lung cancer screening. IEEE Rev Biomed Eng 2009;2:136–146.

82. Robinson C, Halligan S, Taylor SA, Mallett S, Altman DG. CT colonography: a systematic review of standard of reporting for studies of computer-aided detection. Radiology 2008; 246(2):426–433.

83. Halligan S, Altman DG, Mallett S, et al. Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. Gastroenterology 2006;131(6):1690–1699.

84. Bell RM, Koren Y. Lessons from the Netflix prize challenge. SIGKDD Explor Newslett 2007;9(2):1931–1945.

85. Bell RM, Koren Y. Scalable collaborative filtering with jointly derived neighborhood interpolation weights. Seventh IEEE International Conference on Data Mining, 2007; 43–52.