ELSEVIER

European Journal of Radiology 31 (1997) 97–109

EJR
EUROPEAN JOURNAL OF RADIOLOGY

www.elsevier.nl/locate/ejrad

THEME

Computer-aided diagnosis in radiology: potential and pitfalls

Kunio Doi *, Heber MacMahon, Shigehiko Katsuragawa, Robert M. Nishikawa, Yulei Jiang

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

Received 18 January 1999; accepted 19 January 1999

Abstract

Computer-aided diagnosis (CAD) may be defined as a diagnosis made by a physician who takes into account the computer output as a second opinion. The purpose of CAD is to improve the diagnostic accuracy and the consistency of the radiologists' image interpretation. This article is to provide a brief overview of some of CAD schemes for detection and differential diagnosis of pulmonary nodules and interstitial opacities in chest radiographs as well as clustered micro-calcifications and masses in mammograms. ROC analysis clearly indicated that the radiologists' performances were significantly improved when the computer output was available. An intelligent CAD workstation was developed for detection of breast lesions in mammograms. Results obtained from the first 10 000 cases indicated the potential of CAD in detecting approximately one-half of 'missed' breast cancer.

Keywords: Computer-aided; Radiology; Technology

1. Introduction

Over the last decade or so, many investigators have carried out basic studies and clinical applications toward the development of modern computerized schemes for detection and characterization of lesions in radiologic images, based on computer vision and artificial intelligence. These methods and techniques are generally called computer-aided diagnosis (CAD) schemes. The development of CAD has now reached a new phase, since the first commercial unit for detection of breast lesions in mammograms by R2 Technology, Inc., Los Altos, CA, was approved in June, 1998, by the Food and Drug Administration (FDA) for marketing and sale for clinical use in the United States. It is likely that this will expedite clinical applications of the CAD schemes and also further the development of various CAD schemes in different diagnostic examinations. Our purpose in this article is to provide a brief overview of some of the current CAD schemes developed at the University of Chicago for chest radiography and mammography, and to discuss some issues related to the future potential and pitfalls of CAD.

2. Basic concept of computer-aided diagnosis

CAD may be defined as a diagnosis made by a physician who takes into account the results of the computer output as a 'second opinion' [1–6]. In radiology, the computer output is derived from quantitative analysis of diagnostic images. It is important to note that the computer is used only as a tool to provide additional information to clinicians, who will make the final decision as to the diagnosis of a patient. Therefore, the basic concept of CAD is clearly different from that of 'automated diagnosis,' which had been investigated in the 1960s and 1970s. The purpose of CAD in radiology is to improve the diagnostic accuracy as well as the consistency of radiologists' image interpretation by using the computer output as a guide. The computer output can be very helpful because a radiologist's diagnosis is made based on subjective judgement and because radiologists tend to miss lesions such as lung

* Corresponding author. Tel.: +773-702-6954; fax: +773-702-0371.

E-mail address: k-doi@uchicago.edu (K. Doi)

0720-048X/99/$ - see front matter
PII: S0720-048X(99)00016-9

nodules in chest radiographs, and microcalcifications and masses in mammograms. In addition, variations in diagnosis, such as inter-observer and intra-observer variations, can be large.

Usually, two types of general approaches are employed in computerized schemes for CAD. One is to find the location of lesions such as lung nodules in chest images by searching for isolated abnormal patterns with a computer. Another is to quantify the image features of normal and/or abnormal patterns, such as lung texture related to interstitial disease in chest images and vessel sizes related to stenotic lesions in angiograms.

Computerized schemes for CAD generally include three basic components which are based on three different technologies. The first component is image processing for enhancement and extraction of lesions. It is important to note that the image processing involved in CAD schemes is aimed at facilitating the computer, rather than the human observer, to pick up the initial candidates of lesions and suspicious patterns. Various image-processing techniques have been employed for different types of lesions. Some of the commonly used techniques include filtering based on Fourier analysis, wavelet transform, morphological filtering, the difference image technique, and artificial neural networks (ANNs).

The second component is the quantitation of image features such as the size, contrast, and shape of the candidates selected in the first step. It is possible to define numerous features based on some mathematical formula that may not be easily understood by the human observer. However, it is generally useful to define, at least at the initial phase of CAD development, image features that have already been recognized and described subjectively by radiologists. This is because radiologists' knowledge is based on their observations of numerous cases over the years, and their diagnostic accuracy is generally very high and reliable. One of the most important factors in the development of CAD schemes is to find unique features that can distinguish reliably between a lesion and other, normal anatomic structures.

The third component is data processing for distinction between normal and abnormal patterns, based on the features obtained in the second step. A simple and common approach employed in this step is a rule-based method, which may be established based on the understanding of lesions and other normal patterns. Therefore, it is important to note that the rule-based method may provide useful information for improving the CAD schemes. Other techniques used include discriminant analysis, ANN, and the decision-tree method. It is our experience that the combination of the rule-based method with other methods such as ANN tends to produce the best results in CAD schemes.

Because the basic concept of CAD is broad and general, CAD is applicable to all imaging modalities, including conventional projection radiography, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound imaging, and nuclear medicine imaging. In addition, computerized schemes for CAD can be developed for all kinds of examinations in every part of the body, including the skull, chest, abdomen, bone, and vascular system. However, the research subjects investigated for CAD in the past have been limited. The major focus on recent research subjects has been in the field of mammography for early detection of breast cancer. Many CAD schemes are related to detection and characterization of masses and clustered microcalcifications in mammograms [7–14]. In chest radiography [15–23], computerized schemes have been developed for detection of lung nodules, interstitial infiltrates, cardiomegaly, and pneumothoraces. Computerized schemes in angiography [24–26] include the quantitative analysis of stenotic lesions and the determination of pulsatile blood flow rates.

3. Computer-aided diagnosis in chest radiography

It is a difficult task for radiologists to detect some lung lesions, such as nodules. It is well documented that radiologists may miss up to 30% of lung nodules in chest radiographs, because of the camouflaging effect of the normal anatomic background. It is expected, therefore, that the computer prompt by indicating potential sites of nodules would improve their detection accuracy. The computerized scheme for detection of nodules [15,17] is shown in Fig. 1. The digital or digitized chest

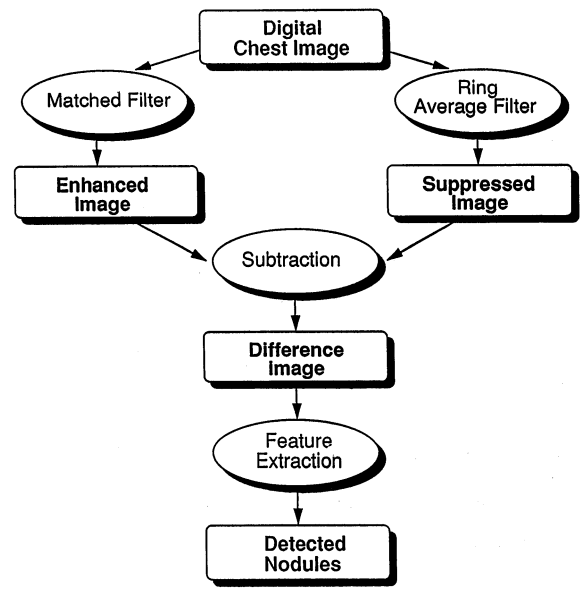


Fig. 1. Overall scheme for automated computerized detection of lung nodules on chest images.

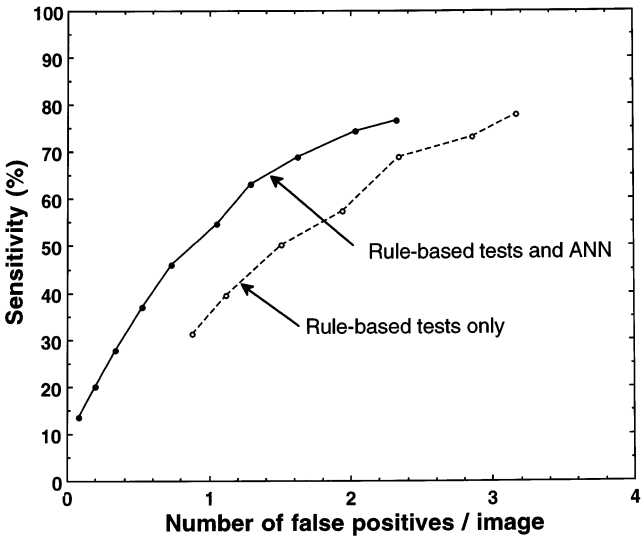


Fig. 2. FROC curves for computerized detection of lung nodules without and with the use of artificial neural network (ANN).

image is first processed by two different filtering operations, namely, one for enhancement of nodules by a matched filter and the other for suppression by a ring average filter. The difference image is then obtained by subtracting of the nodule-suppressed image from the nodule-enhanced image. With this difference image technique, the majority of background structures in chest images is suppressed, and thus initial candidates of nodules can be identified by application of a thresholding technique to the difference image.

A number of image features such as the size, contrast, and shape are determined for all of the initial nodule candidates. These features are applied for distinguishing between nodules and other normal structures such as ribs, pulmonary vessels, and their crossings, in order to remove some of the potential false positive detections by the computer. Because some features of nodules are often different from those of some normal structures, it is possible to apply a rule-based classification method initially to distinguish some normal structures from nodules [27]. For example, whereas nodules tend to be circular, ribs and rib crossings tend to be rectangular. In addition, the size and shape of rib-rib crossings change relatively rapidly as the threshold level applied to the difference image changes, whereas those of nodules change relatively slowly. Because it is not possible to remove all of the false positives by a rule-based classification method, an ANN is used as a second classification method for removal of some remaining candidates which are due to normal structures in chest images.

Fig. 2 shows the free-response receiver operating characteristic (FROC) curves which indicate the relationship between the sensitivity in detecting nodules and the number of false positives per image. It is apparent that the FROC curve can be improved by use of the ANN in removing some false positives [17]. The current level of the performance of our CAD scheme for nodule detection is approximately 70% sensitivity and 1.7 false positives per image. Fig. 3 illustrates a chest image with a nodule, in which the computer output, marked by two arrows, indicates a correct detection of the nodule and a false positive. It is important for any CAD scheme to reduce the number of false positives as much as possible. However, it is of interest to note that the majority of false positives of lung nodules detected by computer are different from those detected by radiologists [28], and thus radiologists can generally disregard these 'obvious' false positives identified by computer.

One important question regarding the value of CAD is whether radiologists can really improve their performance if the computer output such as that shown in Fig. 3 is presented to them for their interpretation. The answer to this question is clearly demonstrated by the results illustrated in Figs 4 and 5, which were obtained from observer performance studies [16]. It is apparent in Fig. 4 that the receiver operating characteristic (ROC) curve in detecting nodules by radiologists was improved significantly ($P = 0.001$) when the computer output was presented to them. Fig. 5 shows the Az values (Az is the area under the ROC curve and a measure of the observer's detection performance). For all sixteen radiologists who participated in this observer study, their performances were improved when they used the computer output, as indicated by the gains in Az values. Without the computer output, the average

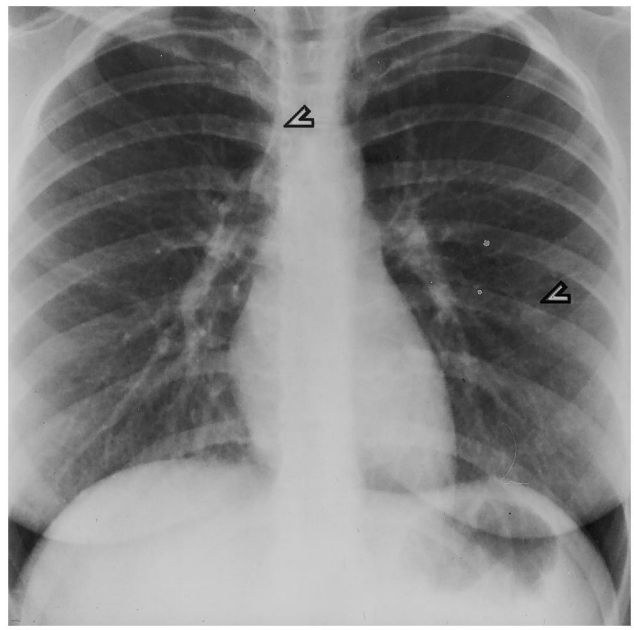


Fig. 3. Illustration of the computer output (arrows) on chest image with a lung nodule.

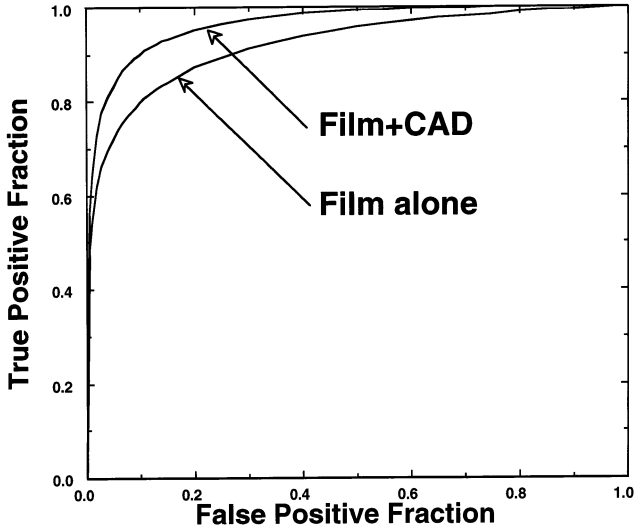


Fig. 4. Comparison of ROC curves for radiologists' detection of lung nodules on chest images without and with computer output.

performance level of eight radiologists, including two chest radiologists and six general radiologists, was greater than that for eight residents, which is understandable. However, the average performance level of the residents, when the computer output was available as CAD, was comparable to that of the other eight radiologists without the computer output. In addition, the difference in the average performance level between residents and other radiologists was less with CAD than without CAD. These results seem to indicate that CAD can improve both the accuracy and the consistency of the radiologic diagnosis. A very similar result was found in a large-scale observer test [29] which was conducted as a real-time ROC study on the detection of lung nodules by use of CRT monitors at the Scientific Exhibit of the Radiological Society of North America (RSNA) meeting in Chicago, November, 1996. More than 100 radiologists participated in that study [29].

We applied our CAD algorithm to examine chest images of 95 lung cancer cases in a mass screening

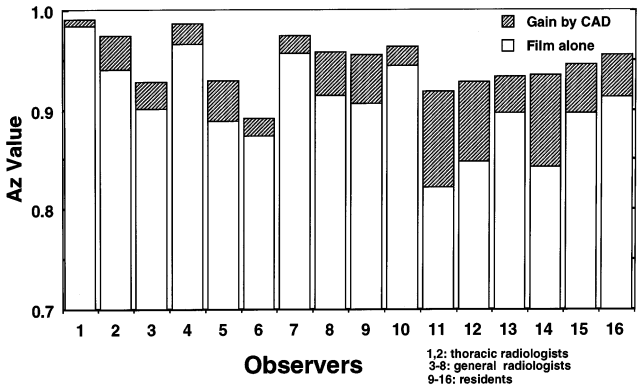


Fig. 5. Az values (areas under ROC curve) obtained by 16 observers without and with computer output.

program in Yamaguchi, Japan [30]. The results are shown in Table 1. Forty-five cases were detected correctly by both mass screening and computer. Twenty-four cases were missed by both screening and computer, and thus, in these cases, detection is obviously very difficult at present. Eleven cases were not detected by computer, but were detected by mass screening. As long as the computer result would have been used as a second opinion, these cancer cases would probably have been detected even with CAD. An important result in Table 1 is the fact that 15 lung cancer cases were detected by computer, but were missed in mass screening. Therefore, if CAD had been implemented in this mass screening, some of these missed cancer cases might have been detected at an earlier stage.

Once a lung nodule in a chest radiograph is detected by the radiologist, the next question is whether the nodule is benign or malignant. A computerized scheme is being developed for determining the likelihood of malignancy of a detected non-calcified lung nodule in chest images [31]. Output from the computer that indicates the likelihood of malignancy would help radiologists to distinguish between benign and malignant lesions, thus improving patient management, for example, by avoiding unnecessary CT examinations on patients with benign lesions. In such a CAD scheme, the detected lung nodules are subjected to quantitative analysis of a number of objective image features which would be correlated with subjective features such as size, contrast, speculation, and uniformity. The likelihood of malignancy is then determined by the output of an ANN which has been trained by use of objective image features, as input to the ANN, for many known cases with benign and malignant nodules. An ROC curve for the computerized scheme for distinction between benign and malignant nodules is shown in Fig. 6. It is apparent in Fig. 6 that the performance level indicated by the ROC curve is much greater for the computerized scheme than for radiologists. Another observer study was performed in which it was ascertained whether radiologists can utilize the ANN output to improve their performance. The ROC curve for radiologists with ANN was indeed greater than that without the ANN.

Another important abnormality in chest radiographs is that of interstitial opacities, which are also considered difficult for radiologists' interpretation. This difficulty is partly related to the fact that the patterns of interstitial opacities are ill-defined and complex. In addition, their variation is complicated. Therefore, the quantitative analysis of opacities based on physical measures may be useful for radiologists' interpretation. The computerized scheme for quantitative texture analysis of interstitial opacities based on Fourier analysis [18] is shown in Fig. 7. A large number (200–600) of regions of interest (ROIs) are initially selected automatically over periph-

Table 1
Distribution of the numbers of cases for true positives and false negatives of 95 lung cancer patients in mass screening in Yamaguchi, Japan, and results by computer

	True positives in mass screening	False negatives in mass screening	Total
True positives by computer	45 (47%)	15 (16%)	60 (63%)
False negatives by computer	11 (12%)	24 (25%)	35 (37%)
Total	56 (59%)	39 (41%)	95 (100%)

eral lung fields. ROIs with sharp edges, such as ribs, are removed from the analysis. The non-uniform background trend in ROIs, which may be due to variation of chest wall thickness and to other normal structures, is corrected by use of a two-dimensional fitting technique. The power spectrum of the infiltrate pattern is obtained by Fourier transformation and filtering with a visual system response. Finally, the rms pixel variation and the first moment of the power spectrum are determined as two texture measures, which are related to the magnitude and coarseness (or fineness), respectively, of the infiltrate patterns.

Fig. 8 shows four chest images which include one normal lung and three abnormal lungs with three different types of interstitial opacities, i.e. with nodular, reticular, and reticulonodular (or honeycomb) patterns. It is generally difficult to distinguish these different types of patterns with confidence, especially subtle abnormalities at an early stage. The corresponding texture measures on these chest images are shown in Fig. 9. It should be noted that, in Fig. 9, texture measures for these four lungs are separated, and thus these objective measures may be helpful to the radiologists' interpretation. The large circle indicates the range of texture measures for normal lungs, which corresponds to the average $\pm$ one standard deviation. Many small open circles, which were obtained from the normal lung in Fig. 8, appear to be distributed as expected within and near the large circle. Texture measures for the nodular pattern are shifted to the left from the normal range, indicating a decrease in the first moment of the power spectrum. Texture measures for the reticular pattern are shifted to the upper region, indicating an increase in the rms variation. However, texture measures for the reticulonodular (honeycomb) pattern indicate changes in both the first moment of the power spectrum and the rms variation. It may be noticed that some of the dark circles for this case are located in the normal range. However, this result was considered a correct representation of texture measures of this case, because the lower lung of this patient was abnormal, whereas the upper part was considered normal.

The results of texture analysis can be represented by markers on chest images as shown in Fig. 10. Normal areas obtained by texture analysis are indicated by pluses, whereas abnormal areas with nodular, reticular, or reticulonodular patterns are indicated by circles, squares (not shown in the case of Fig. 10), or hexagons, respectively. The larger these abnormal markers, the greater the severity of the abnormalities involved. Therefore, the distribution of these markers on chest images may help radiologists in their assessment of the extent of interstitial infiltrates on the images. We conducted an observer performance study to examine whether results such as those shown in Fig. 10 would improve the radiologists' performance in distinguishing between normal lungs and abnormal lungs with interstitial infiltrates. ROC curves obtained in this experiment are shown in Fig. 11. It is apparent in that figure that radiologists' performance was improved ($P = 0.0002$) when the computer results were available [32]. One may ask why this improvement was possible despite the fact that the ROC curve obtained for the computerized texture analysis was lower than that for radiologists alone. This improvement was possible because the radiologists' performance is not identical to the computer's performance, and thus they can complement each other. For example, in some cases radiologists may be correct, but the computer may be incorrect, whereas in other cases the computer may be correct but radiolo-

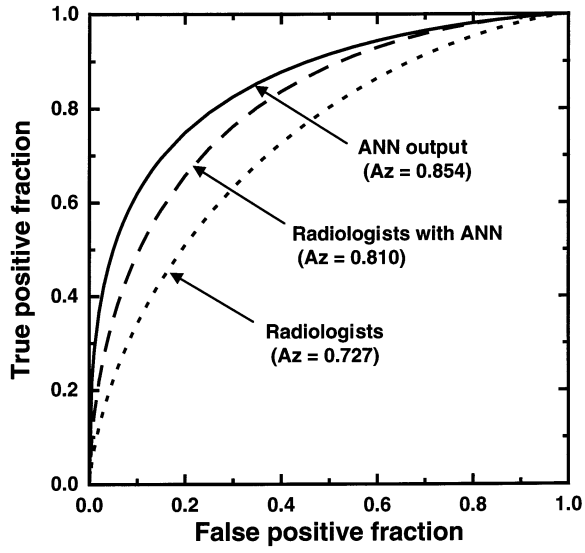


Fig. 6. Comparison of ROC curves in classifying benign and malignant lung nodules on chest images by radiologists alone, radiologists with artificial neural network (ANN) output, and computer output alone.

THEME

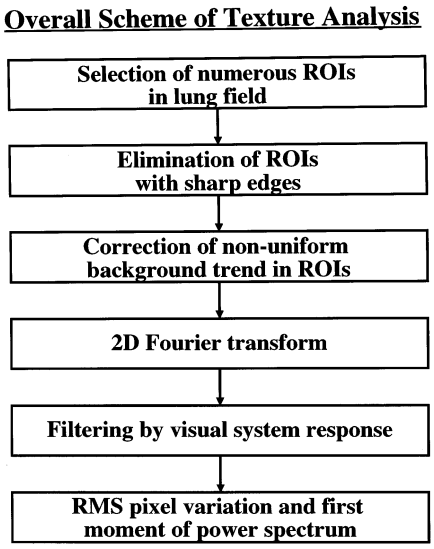


Fig. 7. Overall scheme for automated computerized analysis of lung texture on interstitial infiltrates in chest images.

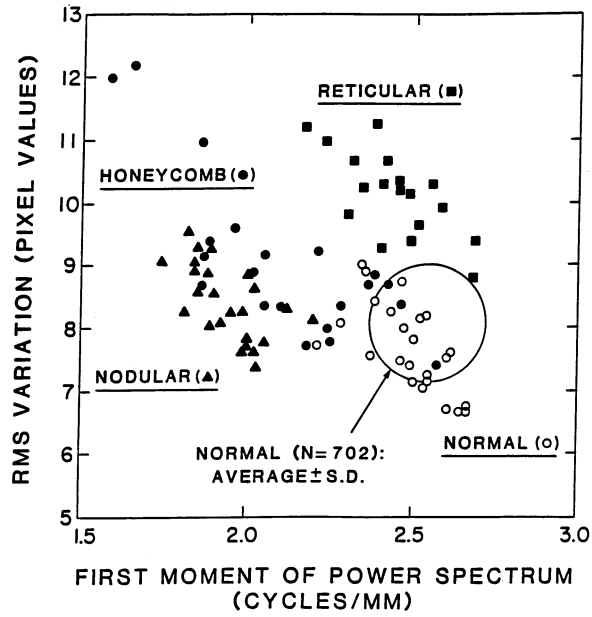


Fig. 9. Distributions of two texture measures obtained from the four chest images illustrated in Fig. 8.

gists may be incorrect. Therefore, a radiologist can correct his/her mistakes occasionally by observing the computer results when he/she had recognized his/her initial mistakes.

Once the abnormality due to interstitial opacities is detected by radiologists, their subsequent task is the differential diagnosis for identification of interstitial disease among many possible diseases. We applied an ANN to determine the likelihood of each of eleven diseases by using a number of clinical parameters and radiologic findings [33], as shown in Fig. 12. All of the clinical parameters and subjective ratings on the radiologic findings were converted to numerical values from 0 to 1.0, which were used as input to the ANN. The output of the ANN also ranged from 0 to 1.0, which corresponds to the likelihood of each disease. The greater the output value of the ANN, the greater the likelihood of the disease corresponding to the output unit. The ANN is initially trained by use of known cases with various diseases, and then the trained ANN is tested with other cases that were not used for training.

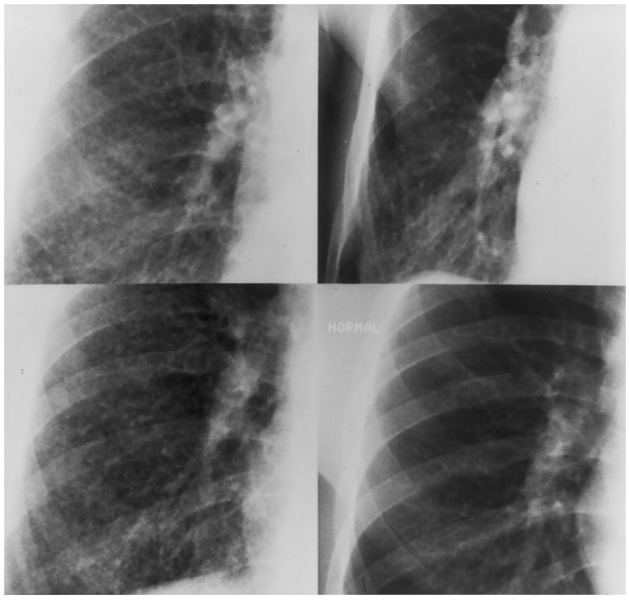


Fig. 8. Illustration of four chest images with one normal lung and three abnormal lungs with nodular, reticular, and reticulonodular (honeycomb) patterns.

The ROC curve of the ANN (solid curve) is compared with that of radiologists (dashed curve), as shown in Fig. 13. It is obvious in the figure that the performance level of the ANN is substantially greater than that achieved by radiologists alone. An important question is, then, whether radiologists' performance can be improved when the output of the ANN as illustrated in Fig. 14 is presented to them. Another ROC curve (dotted) indicates clearly that radiologists' performance in differential diagnosis was improved ($P = 0.0001$) when they used the list of the likelihood of interstitial diseases [34]. However, the ROC curve of radiologists with the ANN output was considerably lower than that of the computer output alone. This result may be due to the lack of familiarity with the ANN output by radiologists at present. When radiologists become familiar with the use of the ANN output, it is expected that the overall performance in the differential diagnosis by radiologists by taking into account the ANN output will be improved further in the future.

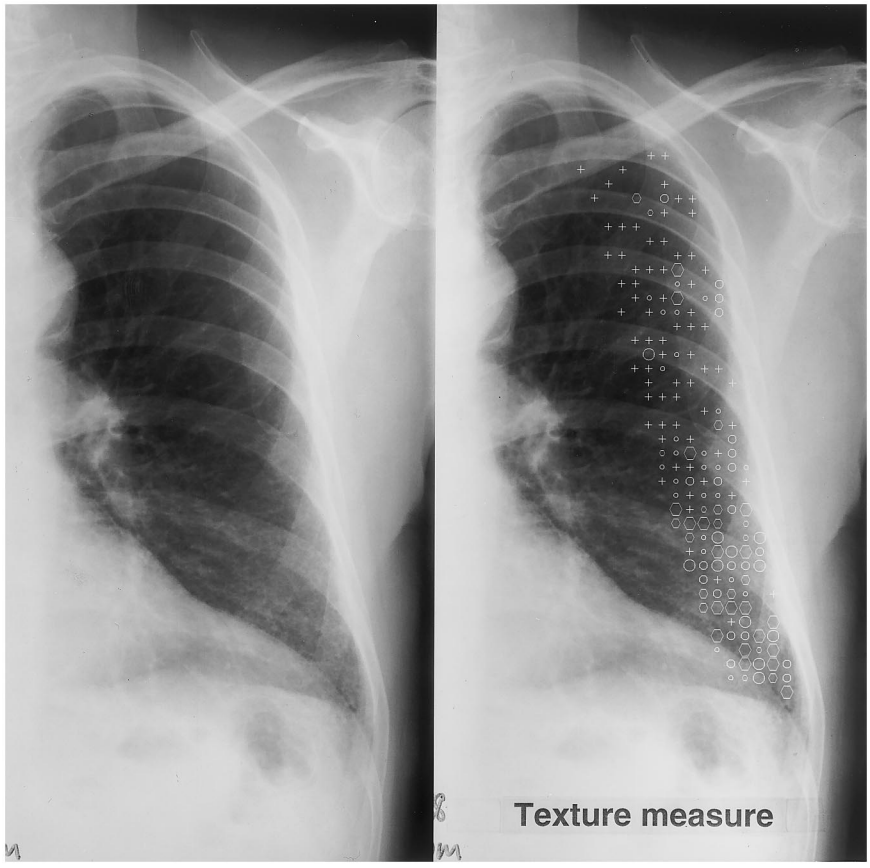


Fig. 10. Illustration of texture measures by using various markers, which were superimposed at numerous locations of ROIs automatically selected on chest images.

4. Computer-aided diagnosis in mammography

Mammography is considered the most reliable method for early detection of breast cancers at present. However, it is difficult for radiologists to detect breast lesions such as clustered microcalcifications and masses on mammograms. It is well known that radiologists may miss 15–30% of these lesions. Therefore, the computer output indicating the potential sites of lesions may be useful to assist radiologists' interpretation of mammograms, especially in mass screening, due to the fact that the majority of cases are normal and only a small fraction are breast cancers. The computerized scheme for detection of clustered microcalcifications [8–12] is shown in Fig. 15, which is similar to the difference image technique used for detection of lung nodules in chest images, as described earlier. However, because microcalcifications are much smaller than lung nodules, a different set of filters is used for identifying the initial candidates for microcalcifications. In addition, different kinds of image features are employed for distinguishing microcalcifications from other, normal patterns and also image noise in mammograms. Because microcalcifications are very small (in the range of 0.1–0.5 mm), it is important to reduce the effect of image noise on the accurate detection of microcalcifications.

An FROC curve for detection of clustered microcalcifications by the computerized scheme is shown in Fig. 16. The current performance level by computer is approximately 85% sensitivity, with a false positive rate of

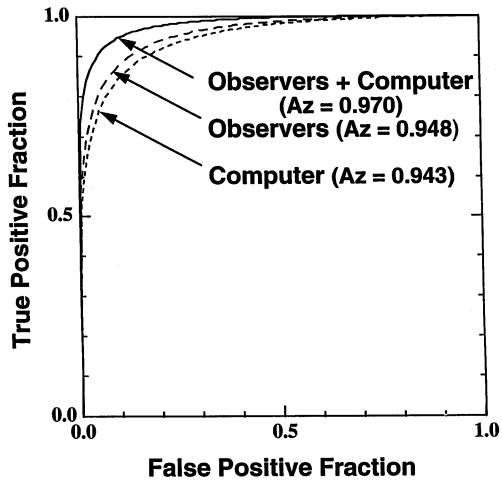


Fig. 11. Comparison of ROC curves for distinction between normal and abnormal lungs with interstitial infiltrates by radiologists alone, radiologists with computer, and the computer output alone.

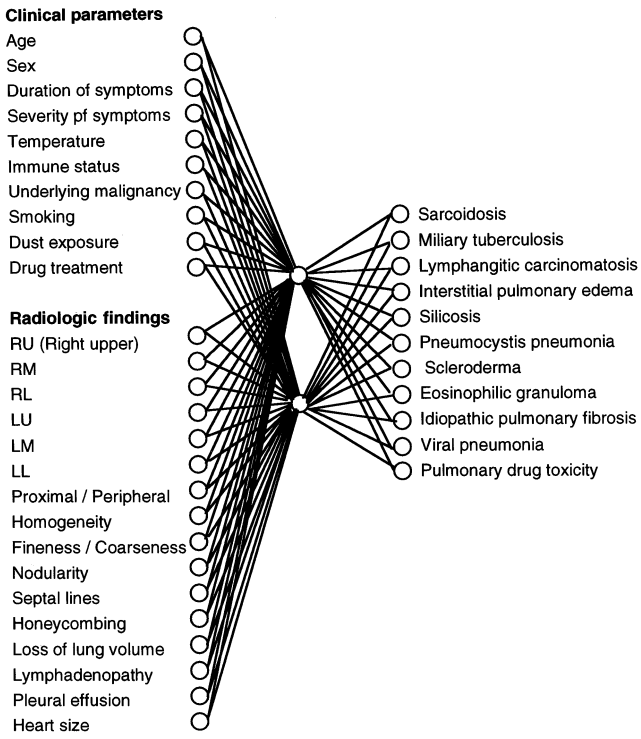


Fig. 12. An artificial neural network applied for differential diagnosis on 11 interstitial diseases (11 output units) based on 10 clinical parameters and 16 radiologic findings (26 input units). For simplicity, only two hidden units are shown.

0.5 per image. Fig. 17 shows a mammogram with subtle clustered microcalcifications, which were correctly detected as indicated by an arrow, without a false positive. An important question is again whether radiologists can take advantage of this type of computer output for improving their detection perfor-

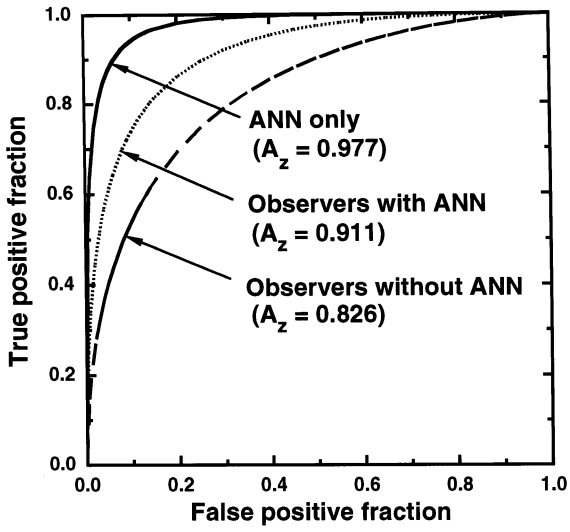
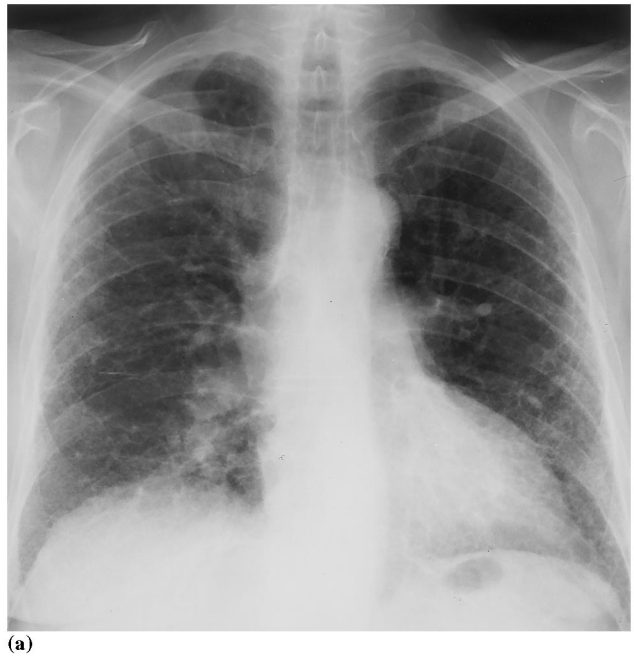
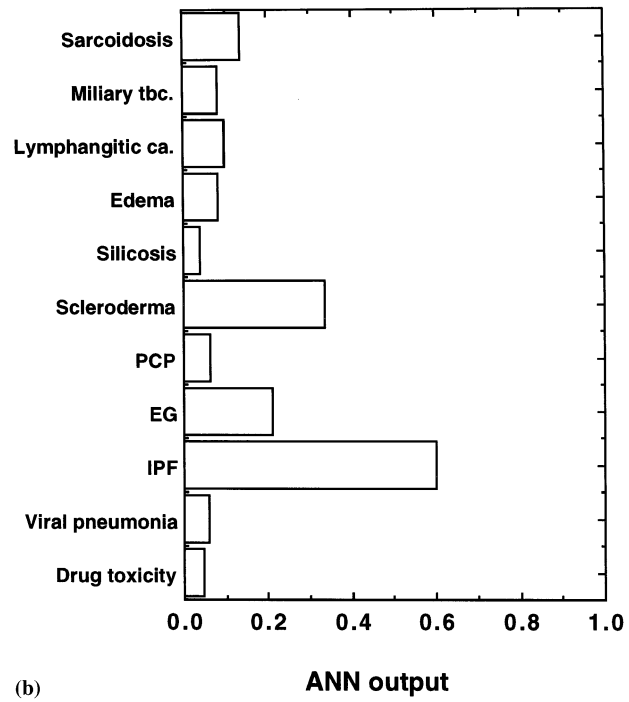


Fig. 13. Comparison of ROC curves for differential diagnosis of 11 interstitial diseases by radiologists alone, radiologists with computer, and the ANN computer output alone.



(a)



(b)

Fig. 14. Chest image (a) for 64-year-old man with idiopathic pulmonary fibrosis, and ANN output (b) presented to observer.

mance. ROC curves for detecting subtle clustered microcalcifications without and with the computer output are shown in Fig. 18. It is apparent that radiologists' performance was again improved significantly ($P = 0.001$) when they used the computer output [9]. At a specificity of 95%, the sensitivity in detecting clustered microcalcifications was improved from 80 to 90%. Another way of looking at these data is that radiologists may miss 20% of subtle clustered microcalcifications;

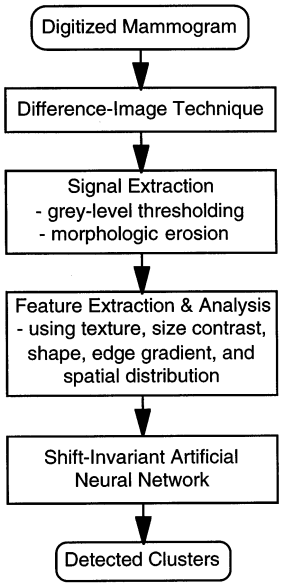


Fig. 15. Overall scheme for automated computerized detection of clustered microcalcifications on mammograms.

however, the CAD scheme may help them to detect 50% of potentially "missed" subtle lesions. Very similar findings from studies on actually missed breast cancer cases have been reported [35,36].

Another important computerized scheme in mammography is the detection of masses, as shown in Fig. 19, where a nonlinear bilateral subtraction technique [7] is employed for identifying the initial candidates of masses in mammograms of the right and left breasts. With this technique, asymmetric mass-like densities are detected by comparison of two mammograms at a number of comparable levels of the corresponding histograms. This nonlinear bilateral subtraction technique was effective in identifying more mass lesions than were identified with the linear subtraction technique. The FROC curve for detection of masses by this scheme is

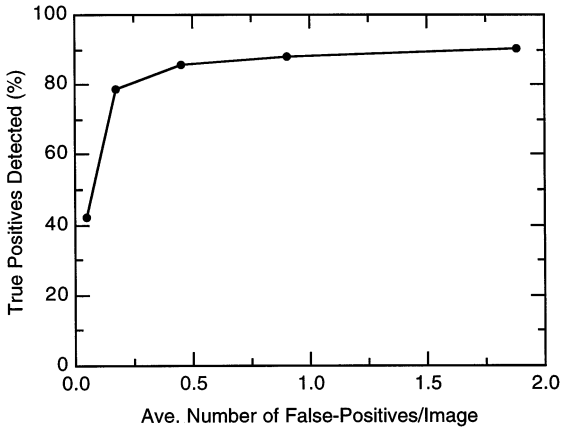


Fig. 16. FROC curve for computerized detection of clustered microcalcifications on mammograms.

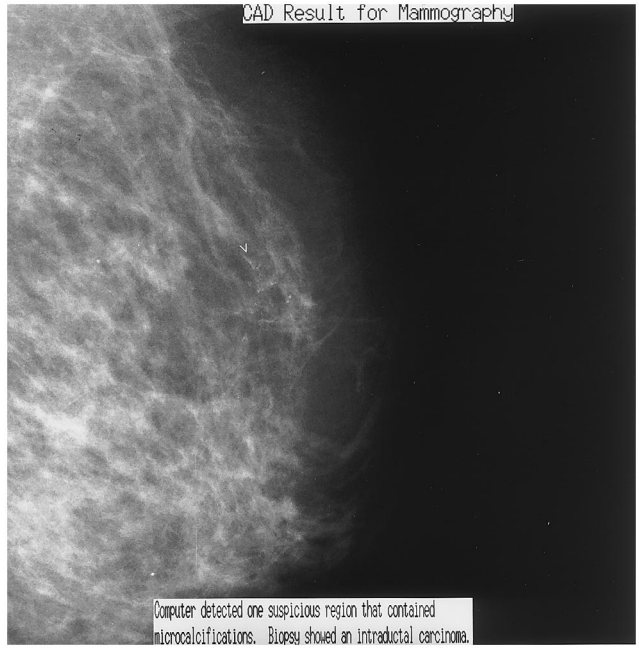


Fig. 17. Illustration of the computer output (arrow) indicating the correct detection of clustered microcalcifications on mammogram.

shown in Fig. 20. The sensitivity is approximately 90% and the number of false positives is 2–3 per image. Fig. 21 shows a pair of mammograms, one of which contains a speculated mass, which was correctly detected by computer. However, there were two false positives on normal breast tissue.

Once a breast lesion is detected by the radiologist, the next question is whether the lesion is benign or malignant. The computerized analysis of these lesions in determining the likelihood of malignancy may be helpful to radiologists for their decision on patient management. This may reduce the number of unneces-

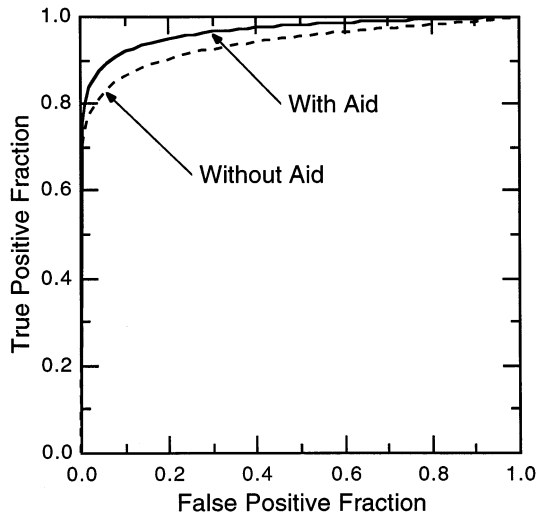


Fig. 18. ROC curves for radiologists' detection of clustered microcalcifications on mammograms without and with computer output.

THEME

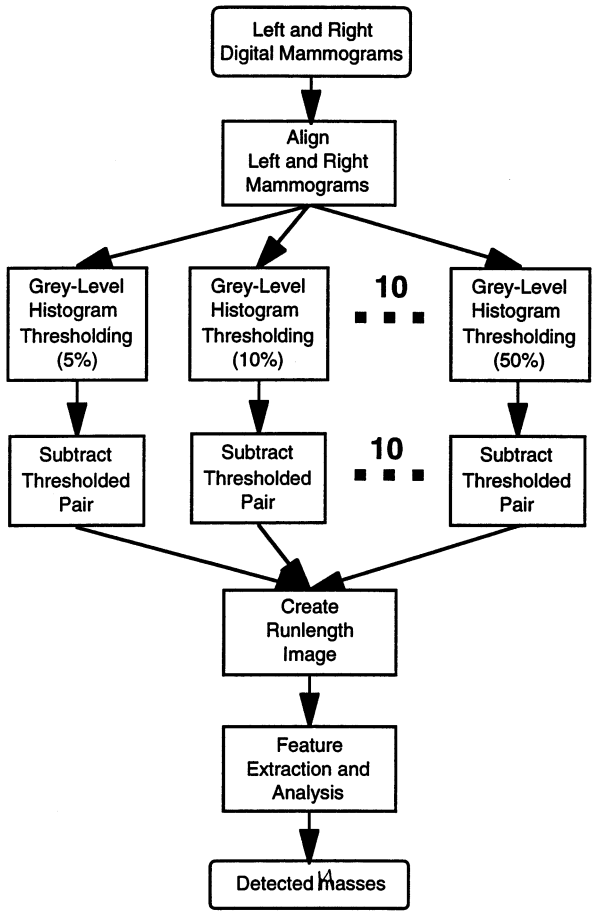


Fig. 19. Overall scheme for automated computerized detection of masses on mammograms.

sary biopsies for patients with benign lesions. A computerized scheme by use of an ANN was developed to merge a number of image features on individual microcalcifications and also the cluster, in order to provide the likelihood of malignancy [13]. Fig. 22 shows clus-

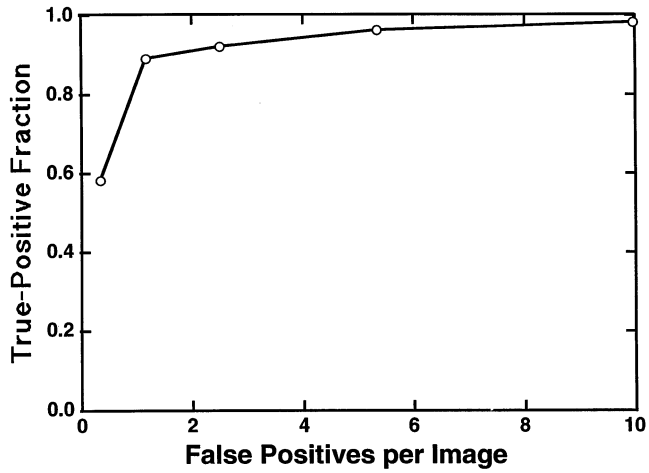


Fig. 20. FROC curve for computerized detection of mammographic masses.

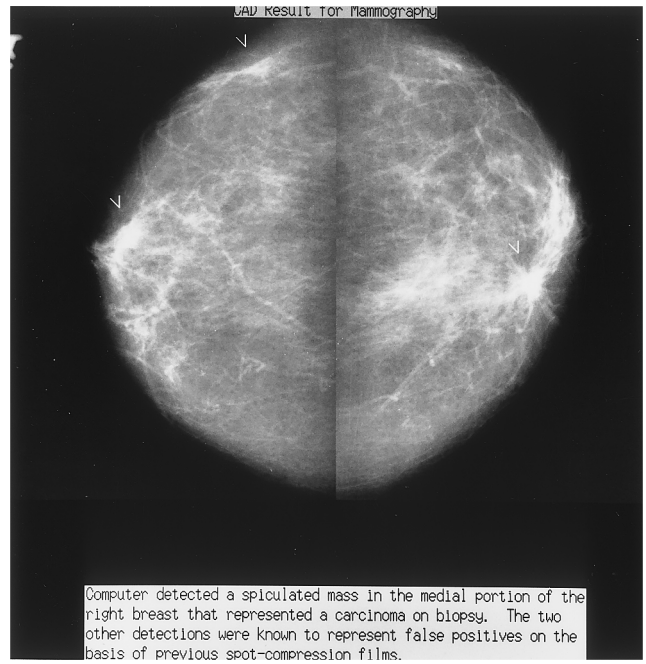


Fig. 21. Illustration of the computer output by three arrows: one indicates correct detection of a speculated lesion, and two are false positives indicating normal breast tissues.

tered microcalcifications of a malignant lesion, where the outline of the cluster was delineated by use of morphological filters and then was subjected to the determination of features such as the size and shape. Observer performance studies were conducted for classifying clustered microcalcifications as malignant or

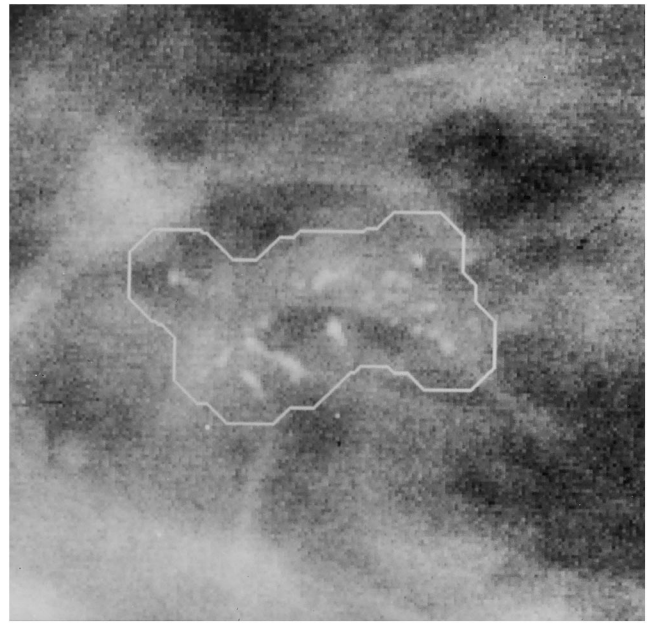


Fig. 22. Illustration of a malignant lesion with clustered microcalcifications and delineation of the outline of the cluster by use of morphologic filters.

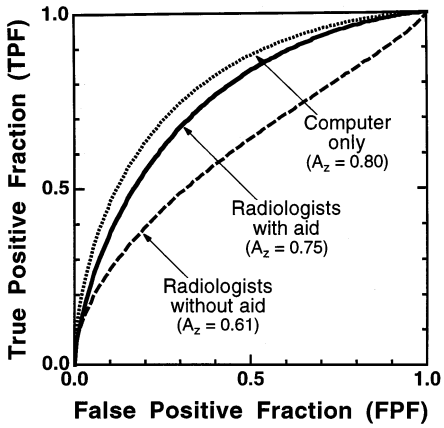


Fig. 23. Comparison of ROC curves for distinction between benign and malignant clustered microcalcifications by radiologists alone, radiologists with computer output, and computer output only.

benign without and with the computer output indicating the likelihood (%) of malignancy. The ROC curves shown in Fig. 23 indicate that the performance level of the computerized scheme was considerably greater than that of radiologists. However, the radiologists' performance was improved significantly when the computer output was used [37]. This result and all other results obtained in our observer studies as described in this article indicate consistently that radiologists were able to use the computer output as a second opinion to improve their diagnostic accuracy.

Based on these findings, we developed an intelligent CAD workstation [4] for detection of masses and clustered microcalcifications, as shown in Fig. 24. The CAD workstation includes a laser film digitizer, a high-speed computer, high-resolution CRT monitors, and a video printer. This workstation was implemented in November, 1994, in the mammography section of our department. Since then, mammograms for more than 16 000 cases have been digitized and analyzed by our computer programs. Among the first 10 000 cases examined, there were 61 women who had developed breast cancers. Forty-seven breast cancers were initially detected in a screening mammogram, whereas 12 cancer cases were considered 'true' negative in screening mammograms because lesions were not visible even in retrospect. Two cases were read as negative, but the cancer was visible in retrospect. Among the 49 mammographically visible cancers, the computer identified 34 (69%) lesions, which included 25 (74%) of 34 masses and nine (60%) of 15 clustered microcalcifications. The sensitivity of detecting these lesions is apparently lower than that obtained by use of our databases. This may be due to the difference between the prospective study with unknown cases and the retrospective study with known cases. In addition, our programs had been frozen about 4 years ago for this prospective study. We believe, however, that the sensitivity can be improved in the future by upgrading of the computer programs with recent development.

Fig. 24. An intelligent CAD workstation for computerized detection of masses and clustered microcalcifications, which includes a laser film digitizer, a high-speed computer, high-resolution monitors, and a video printer.

It is important to note that there were 30 breast cancer cases which had a negative screening examination at least approximately 1 year prior to the correct cancer diagnosis. Among them, 14 cases were considered 'true' negative mammographically, whereas the lesions in 16 cases were visible in retrospect. However, the computer correctly detected nine (56%) of 16 visible lesions. This is an encouraging result, because it indicates the potential of CAD in detecting approximately one-half of 'missed' breast cancers. Fig. 25 shows two pairs of mammograms which were obtained in 1995 and 1996. The computer output pointed to a subtle lesion in the mammogram in 1995 (Fig. 25a), but this was not detected by the radiologist. However, the same lesion was detected correctly in a subsequent mammogram by both the radiologist and the computer (Fig. 25b).

5. Discussion and conclusion

We believe that there is a strong evidence of the potential benefit of CAD in the detection and characterization of some lesions in chest radiography and mammography. However, it is important to be cautious about potential pitfalls associated with the use of the computer output. Advances in science and technology can bring many benefits, but also can be harmful if not used properly. Potential pitfalls of CAD can occur for all four of the possible outcomes, namely, false positives and false negatives, and even with true positives and true negatives indicated by computer.

There has been a general concern that false positives determined by computer may increase the number of unnecessary biopsies in mammographic screening.

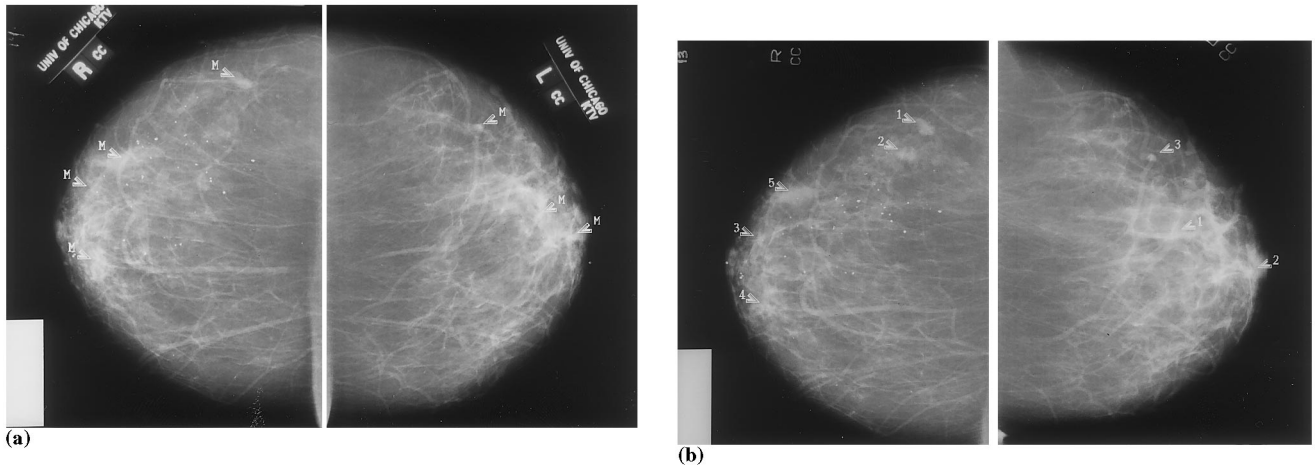


Fig. 25. Comparison of a mammogram (a) in 1995 and a subsequent mammogram (b) in 1996, with the computer output indicating the correct detection of a missed breast cancer in 1995.

However, because many computer false positives are different from radiologists' false positives, which look very similar to 'true' lesions to radiologists, it is unlikely to produce a large increase in biopsy and call-back rate. In fact, two studies [4,36] of CAD in mammography produced similar outcomes. That is, that there was no increase in the call-back rate for additional examinations when CAD was implemented. However, if radiologists are strongly influenced by the computer output, and/or for some other reasons, in determining a threshold level on decision-making such as biopsy versus non-biopsy, there will be a danger of unnecessary biopsies. False negatives identified by computer can cause a problem with missing obvious and detectable lesions if the computer output is trusted excessively, and if radiologists curtail their usual effort in searching for lesions. Therefore, radiologists' expertise and conscientious efforts will remain critically needed. However, radiologists may face difficult situations when they disagree with true positives and true negatives obtained by computer, even if the final decisions are made conscientiously with their best effort. This is obviously a complicated issue that needs to be resolved through experience in the future.

In the meantime, we need to consider what is required for CAD in the near future. Although the performance levels of some CAD schemes are relatively high, it is necessary to improve further the sensitivity and the specificity of detection and classification of many lesions in radiologic images. The ultimate usefulness of CAD is dependent upon the performance levels achievable by these schemes. For example, if both the sensitivity and specificity approach a 100% level, then some simplified tasks may be automated. However, it may take a very long time to realize such a high level of computer performance. A more important and realistic expectation for CAD would be to prove the benefits that are obtainable from the clinical application of current CAD schemes to screening programs for mammography and chest radiography. Results from prospective studies are needed for a further understanding of the potential impact of CAD on the practical aspects of clinical radiology.

In conclusion, CAD is expected to have a major impact on diagnostic radiology in the future. However, successful implementation will require both short-term and long-term development efforts to further improve the accuracy of CAD schemes.

Acknowledgements

Authors are grateful to numerous current and former members in the Rossmann Lab for their contribution to the development of CAD schemes, and also Mrs E. Lanzl for improving the manuscript. This work was supported by the NIH (USPHS Grants CA 60187, CA 64370, CA 62625, AR 42739, HL 52567, T32 CA 09649, RR 11459), US Army Medical Research and Materiel Command (DAMD 17-94-J-4071, DAMD 17-96-1-6058, DAMD 17-96-1-6228, DAMD 17-96-1-6229, DAMD 17-97-1-7202), the American Cancer Society (FRA-390), and The Whitaker Foundation. K. Doi, H. MacMahon, R. M. Nishikawa, and M. L. Giger are shareholders of R2 Technology, Inc., Los Altos, CA. It is the policy of the University of Chicago that investigators disclose actual or potential significant interests that may be affected by research activities.

References

[1] MacMahon H, Doi K, Chan HP, Giger ML, Katsuragawa S, Nakamori N. Computer-aided diagnosis in chest radiology. J Thoracic Imaging 1990;5:67–76.
[2] Doi K, Giger ML, Nishikawa RM, Hoffmann KR, MacMahon H, Schmidt RA, Chua KG. Digital radiography: A useful clinical tool for computer-aided diagnosis by quantitative analysis of radiographic images. Acta Radiologica 1993;34:426–39.

[3] Giger ML, Doi K, MacMahon H, Nishikawa RM, Hoffmann KR, et al. An "intelligent" workstation for computer-aided diagnosis. RadioGraphics 1993;13:647–56.
[4] Nishikawa RM, Haldemann RC, Papaioannou J, Giger ML, Lu P, Schmidt RA, Wolverton DE, Bick U, Doi K. Initial experience with a prototype clinical "intelligent" mammography workstation for computer-aided diagnosis. Proc SPIE 1995;2434:65–71.
[5] Giger ML, MacMahon H. Image processing and computer-aided diagnosis. In: Greenes RA, Bauman R, editors. Radiologic Clinics of North America, vol. 34. Philadelphia: Saunders, 1996:565–96.
[6] Doi K, Giger ML, Nishikawa RM, Schmidt RA. Digital Mammography. Amsterdam: Elsevier Science, 1996.
[7] Yin FF, Giger ML, Doi K, Metz CE, Vyborny CJ, Schmidt RA. Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images. Med Phys 1991;18:955–63.
[8] Chan HP, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM. Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography. Med Phys 1987;14:538–48.
[9] Chan HP, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam KL, Ogura T, Wu Y, MacMahon H. Improvement in radiologists' detection of clustered microcalcifications on mammograms: The potential of computer-aided diagnosis. Invest Radiol 1990;25:1102–10.
[10] Nishikawa RM, Giger ML, Doi K, Vyborny CJ, Schmidt RA. Computer-aided detection of clustered microcalcifications: An improved method for grouping detected signals. Med Phys 1993;20:1661–6.
[11] Yoshida H, Doi K, Nishikawa RM, Giger ML, Schmidt RA. An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms. Acad Radiol 1996;3:621–7.
[12] Zhang W, Doi K, Giger ML, Nishikawa RM, Schmidt RA. An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms. Med Phys 1996;23:595–601.
[13] Jiang Y, Nishikawa RM, Wolverton DE, Metz CE, Giger ML, Schmidt RA, Vyborny CJ, Doi K. Malignant and benign clustered microcalcifications: Automated feature analysis and classification. Radiology 1996;198:671–8.
[14] Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Schmidt RA, Doi K. Automated computerized classification of malignant and benign mass lesions on digitized mammograms. Acad Radiol 1998;5:155–68.
[15] Giger ML, Doi K, MacMahon H. Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields. Med Phys 1988;15:158–66.
[16] Kobayashi T, Xu X-W, 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;199:843–8.
[17] Xu XW, Doi K, Kobayashi T, MacMahon H, Giger ML. Development of an improved CAD scheme for automated detection of lung nodules in digital chest images. Med Phys 1997;24:1395–403.
[18] Katsuragawa S, Doi K, MacMahon H. Image feature analysis and computer-aided diagnosis in digital radiography: Detection and characterization of interstitial lung disease in digital chest radiographs. Med Phys 1988;15:311–9.
[19] Ishida T, Katsuragawa S, Kobayashi T, MacMahon H, Doi K. Computerized analysis of interstitial disease in chest radiographs: Improvement of geometric-pattern feature analysis. Med Phys 1997;24:915–24.
[20] Nakamori N, Doi K, MacMahon H, Sasaki Y, Montner S. Effect of heart size parameters computed from digital chest radiographs on detection of cardiomegaly: Potential usefulness for computer-aided diagnosis. Invest Radiol 1991;26:546–50.
[21] Sanada S, Doi K, MacMahon H. Image feature analysis and computer-aided diagnosis in digital radiography: Automated detection of pneumothorax in chest images. Med Phys 1992;19:1153–60.
[22] Armato SG, Giger ML, MacMahon H. Computerized delineation and analysis of costophrenic angles in digital chest radiographs. Acad Radiol 1998;5:329–35.
[23] Giger ML, Bae KT, MacMahon H. Computerized detection of pulmonary nodules in computed tomography images. Invest Radiol 1994;29:459–65.
[24] Fujita H, Doi K, Fencil LE, Chua KG. Image feature analysis and computer-aided diagnosis in digital radiography. 2. Computerized determination of vessel sizes in digital subtraction angiographic images. Med Phys 1987;14:549–56.
[25] Hoffmann KR, Doi K, Chen SH, Chan HP. Automated tracking and computer reproduction of vessels in DSA images. Invest Radiol 1990;25:1069–75.
[26] Hoffmann KR, Doi K, Fencil LE. Determination of instantaneous and average blood flow rates from digital angiograms using distance-density curves. Invest Radiol 1991;26:207–12.
[27] Matsumoto T, Yoshimura H, Doi K, Giger ML, Kano A, MacMahon H, Abe K, Montner SM. Image feature analysis of false positives produced by an automated computerized scheme for the detection of lung nodules in digital chest radiographs. Invest Radiol 1992;27:587–97.
[28] Yoshimura H, Giger ML, Doi K, MacMahon H, Montner SM. Computerized scheme for the detection of pulmonary nodules: Nonlinear filtering technique. Invest Radiol 1992;27:124–9.
[29] MacMahon H, Engelmann R, Behlen F, Hoffmann KR, Ishida T, Roe C, Metz CE, Doi K. Computer-aided diagnosis of pulmonary nodules: Results of a large scale observer test. Radiology 1999 (in press).
[30] Matsumoto T, Doi K, Nakamura H, Nakanisi T. Potential usefulness of computer-aided diagnosis (CAD) in a mass survey for lung cancer using photofluorographic films. Nippon Acta Radiologica 1992;52:74–6.
[31] Nakamura K, Yoshida H, Engelmann RM, MacMahon H, Ishida T, Doi K. Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules on chest radiographs using artificial neural network (ANN). Radiology 1998;209(P):221–2.
[32] Monnier-Cholley L, MacMahon H, Katsuragawa S, Morishita J, Ishida T, Doi K. Computer aided diagnosis for detection of interstitial infiltrates in chest radiographs: Evaluation by means of ROC analysis. Am J Radiol 1998;171:1651–6.
[33] Ashizawa K, Ishida T, MacMahon H, Vyborny CJ, Katsuragawa S, Doi K. Artificial neural networks in chest radiographs: Application to differential diagnosis of interstitial lung disease. Acad Radiol 1999;6:2–9.
[34] Ashizawa K, MacMahon H, Ishida T, Nakamura K, Vyborny CJ, Katsuragawa S, Doi K. Effect of artificial neural network on radiologists' performance for differential diagnosis of interstitial lung disease on chest radiographs. Am J Radiol 1999;172:1311–4.
[35] Schmidt RA, Nishikawa RM, Osnis RB, Giger ML, Schreibman K, Doi K. Computerized detection of lesions missed by mammography. In: Doi K, Giger ML, Nishikawa RM, Schmidt RA, editors. 3rd International Workshop on Digital Mammography. Amsterdam: Elsevier Science, Excerpta Medica International Congress series, 1996:105–10.
[36] Doi T, Hasegawa A, Hunt B, Marshall J, Rao F, Roehrig J, Romsdahl H, Schneider A, Sharbaugh R, Wang B, Zhang W. Clinical results with the R2 ImageChecker TM Mammographic CAD system. In: Doi K, MacMahon H, Giger ML, Hoffmann KR, editors. Computer Aided Diagnosis. Amsterdam: Elsevier, 1998 (in press)
[37] Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K. Improving breast cancer diagnosis with computer-aided diagnosis. Acad Radiol 1999;6:22–3.