I. INTRODUCTION

Computer-aided diagnosis (CAD) can be defined as a diagnosis made by a radiologist who uses the output from a computerized analysis of radiographic images as a "second opinion" in detecting lesions and in making diagnostic decisions. The final diagnosis is made by the radiologist.

Although mammography is currently the best method for the detection of breast cancer, between 10-30% of women who have breast cancer and undergo mammography have negative
mammograms (1-6). In approximately two-thirds of these false-negative mammograms, the radiologist failed to detect the cancer that was evident retrospectively (4,5,7,8). The missed detections may be due to the subtle nature of the radiographic findings (i.e., low conspicuity of the lesion), poor image quality, eye fatigue or oversight by the radiologists. In addition, it has been suggested that double reading (by two radiologists) may increase sensitivity (9-13). Thus, one aim of CAD is to increase the efficiency and effectiveness of screening procedures by using a computer system, as a "second reader" (like a “spell checker”), to aid the radiologist by indicating locations of suspicious abnormalities in mammograms, leaving the final decision regarding the likelihood of the presence of a cancer and patient management to the radiologist (14). Since mammography is becoming a high volume x-ray procedure routinely interpreted by radiologists and since radiologists do not detect all cancers that are visible on images in retrospect, it is expected that the efficiency and effectiveness of screening could be increased by CAD. The interpretation of screening mammograms lends itself to CAD since it is a repetitive task involving mostly normal images.

If a suspicious region is detected by a radiologist, he or she must then visually extract various radiographic characteristics. Using these features, the radiologist then decides if the abnormality is likely to be malignant or benign, and what course of action should be recommended (i.e., return to screening, return for follow-up or return for biopsy). Many patients are referred for surgical biopsy on the basis of a radiographically detected mass lesion or cluster of microcalcifications. Although general rules for the differentiation between benign and malignant breast lesions exist (1,15), considerable variability occurs in the interpretation of lesions by radiologists with current radiographic techniques (6,16). On average, only 10-20% of masses referred for surgical breast biopsy are actually malignant (1,17-20). Thus, another aim of CAD is to extract and analyze the characteristics of benign and malignant lesions seen on mammography in an objective manner in order to aid the radiologist by increasing diagnostic
accuracy and reducing the numbers of false-positive diagnoses of malignancies; thereby decreasing patient morbidity as well as the number of surgical biopsies performed and their associated complications. It is hoped that CAD will improve the reproducibility of mammographic interpretation. Computerized classification methods are also being extended to ultrasound and magnetic resonance images of the breast.

The presentations of the various detection and classification methods vary depending on the details of the specific methodology that are available and on the database employed with which to evaluate the computerized approach. It is important to note that a comparison of different computerized methods is not possible due to the use of different databases. That is, it cannot be assumed that a computerized scheme that achieves a high sensitivity with one database of mammograms will achieve a similar performance level with another database or with an actual patient population. In addition to the database, the method of evaluation will influence performance and expectations of a specific computerized scheme. Some of the computerized schemes have been tested, or merely demonstrated, only with a few images; whereas others have used moderate-size databases and statistical testing methods. Performance levels, in the latter situations, are usually given for the detection schemes in terms of the sensitivity (true-positive rate) for detection and the number of false-positive detections per image; and for the classification schemes in terms of the sensitivity for classification and the specificity (i.e., one minus the false-positive fraction). Such issues will be discussed later.

II. COMPUTER VISION AND ARTIFICIAL INTELLIGENCE

Computer vision involves having a computer extract from a digital image features that may or may not be otherwise perceived by a human. The development of computer vision schemes requires a priori information about the medical image (e.g., the mammogram) and knowledge of various computer processing techniques and decision analysis methods. The
required a priori knowledge includes the physical imaging properties of the digital image acquisition system and morphological information concerning the abnormality (e.g., mass lesion or cluster of microcalcifications) along with its associated anatomic background. That is, a sufficient database is needed in order to cover the entire range of abnormals and normals. Computer vision techniques include, in general, image processing, image segmentation and feature extraction (21,22).

Once the mammographic image is in digital format, the data are accessible for computer manipulations. Image processing techniques can be employed to enhance features for ease of extraction later or to de-emphasize unwanted features such as background structures. Segmentation of an image involves the separation of the image into regions of similar attributes. Basically, a segmentor only subdivides an image and does not try to recognize the individual segments. Feature extraction entails the recognition of specific characteristics of the individual segments.

Once features can be extracted, relevant ones need to be selected and optimally merged by a classifier. Artificial intelligence techniques include feature selection methods such as step-wise selection and genetic algorithms, and classifiers such as discriminant analysis techniques, rule-based expert systems, and artificial neural networks. Such techniques are used to merge computer-extracted features and/or radiologist-extracted features into diagnostic decision aids.

Currently, screen/film combinations are used as the image recording system for mammographic imaging with investigators in CAD using digitized mammograms almost exclusively. Developments in full-field digital mammography (FFDM) (29), however, have approached clinical usage and thus, yielding the opportunity for CAD on FFDM. Digital image acquisition (film digitization or FFDM) can vary in terms of pixel size and number of quantization levels. In general, when developing a detection scheme, it is important to consider the size of the lesion of interest when choosing pixel size. For example, the detection of
microcalcifications requires a much smaller pixel size (probably around 0.1 mm) than does the task of detecting mass lesions. In addition, the classification of a lesion in question (to be discussed later) may require higher spatial resolution (finer digitization) than that needed for detection. Chan et al. (40,41) have examined the effect of various system parameters, such as pixel size, quantization and data compression, on the detection performance of computerized methods. They demonstrated the trade-off between detection accuracy and data compression rate or digitization resolution.

The availability of high-speed computers and high-resolution film digitizers has been instrumental in the development of advanced computerized detection schemes. It is assumed that over 100 research groups in either academia or industry are developing computerized schemes for the detection of masses and clustered microcalcifications in digital mammograms. Many involve the use of classifiers to distinguish between actual lesions and false-positive detections, or between malignant and benign lesions. Examples of these classifiers in CAD schemes include rule-based methods (48,90), discriminant analysis (69), Baysian methods (68, 85), artificial neural networks (55,70,71), and fuzzy logic (76). The features for input to these classifiers are selected using a variety of techniques including step-wise methods (130) and genetic algorithms (56, 84, 130). Researchers are investigating two types of computer-extracted features: radiologist-used features (e.g., see 124, 126) and higher-order features that may not be as intuitive to radiologists (e.g., see 129).

III. COMPUTER-AIDED DETECTION

The computerized detection of lesions involves having the computer locate suspicious regions, leaving the classification of the lesion totally to the radiologist. It has been reported that between 30 to 50% of breast carcinomas detected mammographically demonstrate clustered microcalcifications (23-25), although about 80% of breast carcinomas reveal microcalcifications
upon microscopic examination (26-29). In addition, studies indicate that 26% of nonpalpable cancers present mammographically as a mass while 18% present both with a mass and microcalcifications (25). Thus, most computerized schemes in mammography are being developed for the detection of mass lesions (77-100) or clustered microcalcifications (42-76, 164).

In order to limit the region of search for lesion detection, the breast region may to be initially segmented from the image. Bick, et. al (30) and Mendez et al. (31), have described methods using computer-defined unexposed and direct-exposure image regions to generate a border around the breast region. Once the breast region is known, preprocessing and search can be contained to within the breast border to yield potential lesion sites. Some methods incorporate an equalization of the gray levels near the periphery of the breast in order to correct for the increased density due to the reduced breast thickness (32,33) Most computer detection methods can be described by a preprocessing stage during which a threshold is applied on each pixel based on a feature such as gray level (34), gradient (90), line orientation (35) or a combination of features (36,37) (e.g., from a feature image). Remaining pixels are then grouped yielding lesion candidates from which various features can be extracted and merged in an attempt to reduce false-positive detections.

Various details of developing computer detection methods can be found in proceedings of the International Workshops on Digital Mammography (101-103).

IV. COMPUTER-AIDED DIAGNOSIS

Once a lesion is detected, a radiologist must make a decision concerning patient management -- that is, should the patient return to screening, have a follow-up or be sent for biopsy? This differs from a purely benign versus malignant determination.
Computerized analysis schemes are being developed to aid in distinguishing between malignant and benign lesions in order to improve both sensitivity and specificity. Such methods may use features extracted either by computer or by radiologists. The benefit of computer-extracted features is the objectivity and reproducibility of the measure of the specific feature. However, radiologists employ many radiographic image features, which they seem to extract and interpret simultaneously and instantaneously. Thus, the development of methods using computer-extracted features requires, besides the determination of which individual features are clinically significant, the computerized means for the extraction of each such feature.

It is important to restate that one of the aims of computerized classification is to reduce the number of benign cases sent for biopsy. Such a reduction will be clinically acceptable only if it does not result in unbiopsied malignant cases, however, since the "cost" of a missed cancer is much greater than misclassification of a benign case. Thus, computer classification schemes should be developed to improve specificity but not at the loss of sensitivity.

Various computerized classification systems are being developed using human-extracted features (104-116) and computer-extracted features (117-130). The benefit of using computer-extracted features is the objectivity and reproducibility of the result. However, radiologists employ many radiographic image features, which they seem to extract and interpret simultaneously and instantaneously. The development of methods using computer-extracted features requires initial determination of which individual features are clinically significant, prior to devising means for the extraction of each such feature. During the past ten years, large growth has been seen in the development of computerized classification methods for lesions on breast images. Most methods are evaluated in terms of their ability to distinguish between malignant and benign lesions, however, a few have been evaluated in terms of patient management (i.e., return to screening vs. biopsy).
Ultrasound and magnetic resonance images of the breasts are also being subjected to computerized analysis to help distinguish between cancerous and non-cancerous lesions (131-136 and 138-140). Breast sonography is used as an important adjunct to diagnostic mammography and is typically performed to evaluate palpable and mammographically identified masses in order to determine their cystic vs. solid nature. The accuracy of ultrasound has been reported to be 96-100% in the diagnosis of simple benign cysts. Masses so characterized do not require further evaluation. Breast sonography, however, has not been routinely used to distinguish benign from malignant solid masses because of the considerable overlap in their sonographic appearances. Magnetic resonance imaging (MRI) may aid in the diagnosis of breast cancer as a complementary technique to mammography. MRI has the benefit of yielding three-dimensional information of the breast. Contrast-enhanced MRI of the breast is used to indicate differences between lesions and normal tissue due to the increased vascularity and capillary permeability of tumors. However, some benign lesions are also enhanced with contrast-enhanced MRI, and thus, the specificity of the technique is limited. Mussurakis et al. (137) reported that significant variability exists in the assessment of lesions by humans using MR images. With MRI, both temporal and spatial features of lesions may be useful in the discrimination task.

V. COMPUTERIZED BREAST CANCER RISK ASSESSMENT

Breast cancer risk assessment is expected to aid in appropriate surveillance plans that may include enhanced screening for women at increased risk of breast cancer. Computerized methods are being developed to relate characteristics of mammographic parenchymal patterns to breast cancer risk. Mammographic density has been shown to be strongly related to risk of breast cancer (93). Computerized analysis of mammographic parenchymal patterns may provide an objective and quantitative characterization and classification of these patterns (141-148).
VI. CLINICAL EXPERIENCES AND COMMERCIAL SYSTEMS

The ultimate evaluation of CAD is not how well the computerized method performs in its task, but rather how well the human performs when the computer output is used as an aid. Observer studies have shown that radiologists' performance increased when using a computer output as an aid -- in detection (36,45) and in diagnosis (127). It is important to note though that the type of observer may affect the reported effect of CAD. One needs to determine and report the amount of experience of the observers and their current mammography reading load. In addition, it is important to consider the effect of disease prevalence on the observer study. For example, in screening mammography, typically less than 1% of the cases will result in cancer.

One method of evaluating a CAD method is to determine its performance on a database of missed lesions. Schmidt et al. (158) conducted a retrospective analysis using computer analyses of a series of mammographic lesions missed in routine clinical practice. With a database of 69 missed cases, the computer found 50% of the lesions with a false positive detection rate of approximately 1.3 per image total for both microcalcification cluster and mass detection algorithms (with a range of 0 to 11 false detections per image). Similar results from a missed lesion study were found by te Brake et al. (159). In their study, the routine clinical examination had included double reading. Twenty-two of 65 cases were detected at one false positive finding per image.

It should be noted that the use of CAD does not require a "digital" radiology department. Most likely, primary diagnosis will still be made from original mammographic films for many years. The design of a dedicated CAD workstation would incorporate detection schemes for both clustered microcalcifications and masses. Appropriate man-machine interfaces must be developed for the effective and efficient implementation of these CAD schemes. The basic hardware of the workstation could include a laser film digitizer (or an interface to a digital mammographic system), a high-speed computer, a mechanism for CAD implementation and an
optional link to PACS (picture and archival communications systems). Possible mechanisms for CAD reporting include (1) low-resolution image paper printer or monitor with primary diagnosis from original films, (2) high-resolution film printer with primary diagnosis from original films (or possibly the printed film), or (3) multiple high-resolution monitors with primary diagnosis from the CRT screen. The computer output may be presented to the radiologists either by symbols (arrows, circles, etc.) that may be toggled on and off, or by simple text indicating the location of possible lesions in the mammogram. Radiologists would use the computer-reported results as aids in making their final diagnostic decision.

In the clinical setting, CAD methods might be used as a test invoked by the radiologist upon viewing a particular mammographic case or as a routine screening procedure performed on all examinations. The workstation could be configured to allow the radiologist to control the sensitivity and specificity of the computer output. Obviously, a choice of fewer false-positive lesions would be achieved at the cost of a lower number of true-positive lesions, and vice-versa. This trade-off could be adjusted by the radiologist, depending on the nature of the case material and personal preference. For example, a radiologist might choose an output with high sensitivity for examining high-risk patients being screened for cancer, whereas a lower sensitivity and correspondingly lower false-positive rate might be desired for patients at low risk for cancer.

Since November 1994, researchers at The University of Chicago have been analyzing screening mammograms on an "intelligent" workstation for use as a "second opinion" (160,161). This large-scale evaluation of computer-aided diagnosis in mammography in a clinical environment is being conducted with the CAD code “frozen” since 1994, and to date, over 19,000 mammographic screening cases have been analyzed by the computer system. The four standard screening mammograms for a given case are digitized and analyzed by the system. In a prospective study of 10,000 consecutive screening mammograms (using the 1994 version of the detection methods), the computer indicated cancer about one year before it was clinically
diagnosed in approximately 15% of all cancer cases and in 56% of cases where the cancer was visible in retrospect on a clinically negative-read screening mammogram (161). The computer had a false-positive rate of approximately 1.3 false clusters per image and 2.1 false masses per image. It should be noted that on a subset of cases in which the radiologist used the computer output, no change in call back rate occurred.

Hendricks (162) of the University of Nijmegen compared use of the ImageChecker system (R2 Technology, Inc., Los Altos, California) with double reading with two radiologists. They reported no difference between double reading and CAD.

Sittek and Reiser (38) reported on their initial experience with the ImageChecker system at the University of Munich in which 82% of the malignant lesions were detected by the computer. They found that CAD analysis can be implemented into an existing mammography section without disturbing patient handling. They also noted that use of the system prolonged the patient waiting time by approximately 15 minutes, but felt the extra time was “well worth it”. Currently, due to the false positive rate of the ImageChecker, the investigators reported that the computer marks need to be rechecked by an experienced radiologist. They do expect, however, that use of such a CAD system will increase detection for lesser-experienced radiologists.

A large study (163, 165) involving a retrospective analysis of 1083 consecutive cancer cases with 13 institutions and more than 24 radiologists was performed as part of a FDA approval process for the ImageChecker (R2 Technology, Inc., Los Altos, California). The sensitivity for microcalcification cluster detection was 98.3% and that for mass detection was 72% with an average false positive rate (cluster and mass combined) of one per image. A prospective component of their study included an analysis of 14,817 cases with the CAD system. No statistically significant change was observed in the radiologists’ workup rate when the system was used as a screening tool.
Results from these various “clinical” studies are quite encouraging, thus leading to continued research in CAD and the arrival of more commercial systems (e.g., SecondLook from CADx and Scanis System from Trex Medical), which may seek FDA approval.

VI. DISCUSSION and SUMMARY

Overall CAD systems have been shown to 1) detect lesions missed by radiologists, 2) not affect the recall rate in screening mammography, 3) exhibit sufficient robustness to changes in image acquisition and digitization, 4) integrate well into pre-existing mammography sections, and 5) classify lesions similar to expert mammographers.

As further improvements in the performance of computer methods are obtained, further testing of systems in the clinical environment needs to be done. Due to the success of computerized classification in observer performance studies, plans should include the integration of detection and classification methods into more clinical prototypes. In addition, with the advent of direct digital mammography systems, CAD will be directly incorporated into systems and tested. Investigators should also take advantage of the digital image data from mammographic, sonographic, and MR images of the breast, thus, allowing for the integration of image information from multimodalities.

For the present, it is important to appropriately report CAD results so that they are useful to the research community. When presenting performance results of a computerized image analysis method, the following should be clearly described: (a) acquisition of the digital image and image quality, (b) characteristics of the database including method of verifying diagnostic truth, (c) use of the database(s) with respect to development and testing, (d) method of scoring, (e) type of performance measures, and (f) performance as an aid to observers.

The ultimate test of any computerized analysis scheme will be its ability to actually improve radiologists' accuracy when used as an aid. Since, in CAD, the computer is used as a
second opinion and not alone, it may not need to be perfect. It should be noted that a CAD scheme will be useful even at a less-than-perfect sensitivity, especially if the mammographic lesions detected by the computer do not overlap completely with those detected by a radiologist. However, a high sensitivity, at the likely cost of some acceptable number of false positives, is preferred since false negatives in the screening for breast cancer are significantly worse in terms of clinical outcome than are false-positive diagnoses. It is important to note that currently computer analysis schemes are being used as “second readers” and not as primary readers.

Over thirty years after Winsberg et al. (77) first reported on a computerized method for mammograms, CAD has become a recognized and accepted component of future breast imaging as evidenced by its clear presence at the annual meetings of AAPM and RSNA, and by the fact that a commercial CAD system for mammography has had FDA approval for three years and is in routine clinical use. A systematic and gradual introduction of CAD into radiology departments is necessary so that minimum disruptions occur in current reading habits, thereby insuring the acceptability of CAD and optimal diagnostic performance by the radiologist.

The information technology revolution is now changing methods of interpretation of mammographic images and these changes are being enthusiastically embraced by both the medical profession and the public (patient). This progress can be attributed to a variety of reasons including the following:

- Digital image quality has improved considerably.
- Computers are faster, have sufficient storage capabilities, and are able to efficiently accommodate large images over networks.
- Investigators have realized that the development of robust methods requires large databases of clinical images, which have become more feasible to collect.
- There exists more focused attention to current and new computer vision techniques.
• Radiologists/clinicians have recognized the medical necessity of computerized assistance especially in settings such as screening for cancer.

• Investigators are sensitive to the constraints imposed by the end user, e.g., the radiologist.

• Public awareness of computer applications in medical imaging has increased.

This is just the beginning. In the future, potentially all medical images will have some form of computer analysis performed on them in order to benefit the diagnosis.

ACKNOWLEDGEMENTS

The author is grateful to various members of the Kurt Rossmann Laboratories for Radiologic Image Research for useful discussions. This research was supported in parts by USPHS grants CA60187, T32 CA09649, and RR11459, and U.S. Army Medical Research and Materiel Command (DAMD 17-96-6058, 17-97-1-7202, 17-98-1-8194). M. Giger is a shareholder in R2 Technology, Inc. (Los Altos, CA). It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interests which would reasonably appear to be affected by the research activities.

REFERENCES


