AUTOMATED SYSTEM FOR EARLY BREAST CANCER DETECTION IN MAMMOGRAMS

Isaac N. Bankman*, Dong W. Kim*, William A. Christens-Barry*, Irving N. Weinberg#, Olga B. Gatewood#, and William R. Brody#

The Johns Hopkins University
***Applied Physics Laboratory**
#Radiology Department

ABSTRACT

The increasing demand on mammographic screening for early breast cancer detection, and the subtlety of early breast cancer signs on mammograms, suggest an automated image processing system that can serve as a diagnostic aid in radiology clinics. We present a fully automated algorithm for detecting clusters of microcalcifications that are the most common signs of early, potentially curable breast cancer. By using the contour map of the mammogram, the algorithm circumvents some of the difficulties encountered with standard image processing methods. The clinical implementation of an automated instrument based on this algorithm is also discussed.

INTRODUCTION

The most prevalent cause of cancer in women is breast cancer. In the United States, one in nine women develops breast cancer in her lifetime and every year more than 170,000 new cases are diagnosed. The incidence of breast cancer is more than double that of colorectal cancer, the second major type in women. However, breast cancer is not the major cause of cancer deaths in women. Studies have indicated that early diagnosis and treatment may significantly improve the 5-year survival for breast cancer patients [1,2]. The American Cancer Society recommends a baseline mammogram for all women by the age of 40, a mammogram approximately every other year between the ages of 40 and 50, and yearly mammogram screening after the age of 50. It has been shown that these screening tests contribute to earlier diagnosis and treatment of breast cancer and many insurance carriers have agreed to cover these examinations. Because awareness and willingness for prevention of breast cancer is increasing rapidly, it is possible that mammography will soon be one of the highest volume X-ray procedures that radiology clinics use regularly. In the U.S. today, about 35 million women are older than 50 and in the next several years, the female U.S. population above the age of 50 will increase at a higher rate than before, reaching about 40 million in the year 2000. While the volume of mammograms is expected to increase, many hospitals are decreasing the number of radiology trainees due to budgetary cuts. The well-recognized goal of performing mammography on a larger scale is becoming more difficult to attain due to the lack of trained readers. Furthermore, the economic feasibility aspects of mammographic screening require that more than 50 mammograms per machine be interpreted daily. This volume is far beyond the current capacity of most mammography clinics in the U.S.

Besides the volume problem, mammographic screening has also an interpretation reliability problem due to the subtlety of the early signs of breast cancer. The life of a women can be saved only if breast cancer can be detected at a very early stage. Early detection of breast cancer in mammograms is a subtle pattern recognition problem due to the wide variation in the normal breast tissue, the large variety of radiographic findings associated with breast cancer, and the similarity between early breast cancer signs and some normal tissue structures. One of the widely used early mammographic indicators of breast malignancy is the presence of clustered microcalcifications. An individual microcalcification appears as a bright spot that ranges in size from about 0.1mm to 2mm in a mammogram. In common mammographic practice, the presence of three or more microcalcifications in a small region (less than 1 cm^2) is usually accepted as a cluster. The cluster of microcalcifications is a highly sensitive sign and in many cases it is the first and only sign of an early, potentially curable breast carcinoma [3-5].

With increasing pressure on throughput and the subtlety of early breast cancer signs, the possibility of observer error increases. Fatigue from reading excessive numbers of mammograms contributes to an increase in the number of missed breast cancers [6-8]. Experienced radiologists are aware of the human factors that limit

reliability and they generally stop interpreting mammograms after they have read a certain number on the same day. A reliable, computerized system could contribute much needed speed and accuracy to mammogram interpretation by serving as an assistance device for the radiologist. A computer-aided interpretation tool that indicates suspicious structures in mammograms can allow the radiologist to focus rapidly on the relevant parts of the mammogram. Furthermore, with high resolution film digitization and wide dynamic range, it may be feasible to detect lesions that might otherwise be missed by the radiologist due to their small size or low contrast. An automated system that can recognize reliably early signs of breast cancer and work continuously without fatigue will be a valuable asset for any radiology clinic. Such a system can contribute not only to the availability of vital health care but it has a potential for reducing its cost as well.

In the early breast cancer detection problem, the main goal is to miss as few signs as possible on the mammogram; false negatives can delay diagnosis and preclude the possibility of timely intervention to save the life of the patient. At the same time, false positives are also undesirable because they can cause unwarranted biopsy examinations. Since biopsy requires surgery of the breast, false positives should be minimized. Therefore, in breast cancer detection, both sensitivity and specificity are important, with sensitivity bearing a more vital implication.

Several algorithms have been suggested for detecting clusters of microcalcifications [9-12]. An elegant pioneering algorithm was based on preprocessing of the mammogram for enhancing microcalcifications [9]. This algorithm can be adjusted to provide a sensitivity of 95% or more but introduces 5 or more false positive clusters per mammogram at these sensitivity levels. As stated in the conclusions of [9], to reduce the number of false positives better signal extraction techniques are necessary. A large number of false positives per mammogram that need to be ruled out by the radiologist would cause an undesirable burden in a busy radiology clinic.

In other algorithms such as [10], adequate detection relies on the human operator who has to set manually 10 acceptance thresholds that may vary for different mammograms. Although the image processing and pattern recognition aspects of these algorithms may be effective, they are not directly applicable in a clinical setting due to the human supervision that they require.

Algorithms that use local thresholds derived from the local distribution of intensity values on the mammogram such as [11] rely on the existence of bimodal distributions in local analysis windows where microcalcifications (signal) and normal tissue (noise) form two distinct Gaussian modes. In most cases, the intensity distribution within analysis windows is unimodal and the detection thresholds of this approach are difficult to determine.

A recent approach [12] used clinical information such as age, relatives with breast cancer, biopsy history, breast size, and breast density, combined with shape measurements of microcalcifications using an expert system. This approach yielded 72% accuracy in identifying clusters of microcalcifications.

The high false positive rate of algorithms that use image enhancement may be due to the spectral overlap between signal and noise in the breast cancer detection problem. Because microcalcifications are similar to other small structures in normal tissue as well as small film artifacts, the spatial frequency content of microcalcifications overlaps considerably with that of some normal tissue structures and that of film artifacts. Image enhancement is essentially band-pass filtering in frequency domain and when the spectra of signal and noise overlap to a large extent, the pass band enhances both signal and similar noise components giving rise to false positives. Too little enhancement can preclude the detection of some microcalcifications while too much enhancement can increase significantly the amplitude of small background structures and produce a large number of false detections. The best compromise may change from image to image and can be difficult to determine. Especially when a single enhancement filter is used to enhance all mammograms, poor detection results can be obtained in many cases. This is due to the fact that both microcalcifications and normal tissue structures exhibit a large variability in size and shape in different mammograms. Consequently, the spectra of signal and noise can vary significantly across mammograms. The theoretically optimal approach that has not been applied to mammogram analysis, is to use Wiener filtering [13] that maximizes the correct detection rate. This approach would provide the best band-pass filter for each mammogram, based on prior knowledge of signal and noise spectra. However, two concerns are valid about this approach in the breast cancer detection problem. First, the need for prior knowledge of signal and noise spectra implies a relatively high human guidance for each mammogram where segments of signal and noise have to be indicated to the algorithm. Second, the high level of overlap between signal and noise within the same mammogram undermines the performance of all band-pass filtering approaches including the Wiener filter.

Furthermore, enhancement may introduce an additional difficulty in the development of an appropriate algorithm due to the modification that filtering imparts to the data. The goal of breast cancer detection algorithms is to approximate as closely as possible the recognition performance of experienced radiologists possibly using confirmation by a biopsy examination. Therefore, the target clusters are indicated by radiologists who also provide guidance on the related detection criteria. The interaction with experienced radiologists is essential for the development of a reliable breast cancer detection algorithm. When the image is filtered, in many cases the data used by the algorithm can be considerably different than the data used in visual radiographic interpretation. In such cases the detection criteria and suggestions of the radiologist may not be directly applicable to the algorithm, and consequently the accordance between visual and automated detection decreases.

From the information theory point of view, if the mammogram is digitized appropriately, all the information needed to detect microcalcifications is present in the raw image. Enhancement is an attempt to eliminate irrelevant and obscuring information and to transform the relevant information for more convenient detection. Since all the information needed is in the raw data, it is possible that algorithms that can access the relevant information without enhancement can be developed.

In most available algorithms for breast cancer recognition, the detection is performed by comparing the amplitude of the signal, i.e. the local intensity of the mammogram to a threshold. In difficult pattern recognition problems where the signal and noise are similar in spectral content as well as in amplitude, successful detection has been achieved by extracting relevant features from the data [14] while detection algorithms based only on amplitude performed poorly [15]. Especially when the goal is to approximate the visual interpretation of the data, features that reflect the visual cues convey the most relevant and effective information. Similarly, in mammogram analysis the visual recognition criteria developed by expert radiologists across many years can guide the development of an effective algorithm by suggesting features that characterize microcalcifications. An additional advantage of features that represent visual cues is that they provide a set of parameters that can be easily interpreted. This allows a more effective interaction with radiologists and gives the algorithm a potentially higher degree of acceptance in the radiology community.

In some of the available algorithms for breast cancer recognition, estimates of the local intensity gradient are used for detection because microcalcifications have a relatively higher intensity with respect to their immediate surroundings. This is done by comparing the pixel values within a small square kernel about the size of a microcalcification, to the pixel values outside the kernel. Because square kernels do not match the shapes of microcalcifications adequately, these estimates of local gradients can be misleading. In fact any measurement for characterizing microcalcifications may be inadequate if it is made by observing the interior of a kernel of preset arbitrary shape and size.

Based on the considerations mentioned above, we set the following specifications for the design of a new algorithm:
1. Operation on raw data without enhancement.
2. Use of features representing visual mammogram interpretation criteria.
3. Operation without preset analysis kernels.
4. Operation without assumptions about the statistical distribution of parameters.
5. Completely automated operation without human intervention.

The algorithm that we developed satisfies these specifications and circumvents some of the difficulties encountered in other algorithms.

DATA

The data were obtained by digitizing 9 mammograms from different patients diagnosed to have cancer by radiographic examination as well as biopsy. Each mammogram was annotated by an experienced radiologist who indicated the locations of all clusters of microcalcifications in the mammograms. A total of 13 clusters were annotated.

Mammograms were backlit using a uniform source light box and digitized in overlapping segments of 25.6 mm height by 38.4 mm width. Segments were overlapped by about 20% in each dimension, eliminating the

possibility that a microcalcification might appear on a segment boundary and escape detection during numerical analysis. Each segment was imaged by a Canon FD 50 macro lens (with extension tube) onto a Sony XC-77ce CCD array camera at a spatial resolution of 50 μm. The illumination intensity was adjusted so that saturation did not occur in any of the signal bearing regions of the mammograms. The data from each segment, consisting of a raster array of 512 by 768 pixels of 8-bit gray scale, were stored on magnetic media for subsequent numerical analysis.

DEVELOPED ALGORITHM

The strategy of the detection algorithm is to view the image as a landscape where elevation corresponds to brightness. In this perspective, microcalcifications appear as prominent peaks that stand out with respect to the local surround. A section of a mammogram that contains a microcalcification cluster is shown in Fig. 1 and the corresponding 3-D plot of the cluster is shown in Fig. 2a. The algorithm starts by forming the contour map of the image. The contour plot obtained in the vicinity of the cluster is shown in Fig. 2b. These contours are iso-intensity contours analogous to iso-elevation contours in cartography and therefore, they are *not* obtained by edge detection and *do not* require local gradient estimates.

When the contours are obtained, the detection algorithm focuses on concentric contours. Each set of concentric contours that represents a peak (an individual microstructure) is analyzed separately. From each peak, the algorithm obtains a sequence of contour areas progressing from the highest contour level in that set (small area) towards low contour levels (larger area). Contour areas that are too small or too big to be part of a microcalcification are not included in the area sequence of a peak. The algorithm is designed to determine the area growth sequence of an individual peak when other peaks are close, by accounting for the merging of contours.

The algorithm computes 5 measurements (features) from the area sequence of peaks:

1) Departure. In visual inspection, microcalcifications are bright structures with a relatively sharp appearance in their visually perceived edge. In the landscape view of a digitized image, the perceived sharpness of a microstructure depends on the departure of that peak from the surrounding background. A microstructure with sharp edges is a peak that departs abruptly from the background while a fuzzy microstructure is a peak that departs very gradually from the surrounding background. The departure feature quantifies the sharpness of a microstructure using the area sequence of that peak. In the area sequence of a peak, an abrupt departure from background is reflected as a sudden change in the rate of change of the area sequence near the base of the peak. This information can be obtained by using the second derivative of the area sequence. In order to obtain a departure value that is insensitive to the size of the microstructure (absolute values of the areas), the algorithm computes the first derivative sequence, and sets the departure to the maximal relative change in the first derivative, in the lower half of the peak. The contour level where the departure is obtained is considered the base of the peak, i.e. the immediate background level.

2) Prominence. This parameter reflects the relative brightness cue that is used in visual inspection. This local contrast information contributes to discriminating microcalcifications from both normal tissue structures and film artifacts. The prominence value is set to the number of contours above the departure level and it is approximately proportional to the brightness difference between the brightest region of the microstructure and the immediate surround at the level of departure from the background.

3) Steepness. In addition to the sharpness at the perceived edge which is reflected by the departure feature, the rate of change of intensity throughout a microstructure is a significant property for visual inspection. Generally, normal breast tissue structures appear globally more diffuse than microcalcifications. Such diffuse structures are represented by peaks that have a gradual increase in height. In the landscape view of the mammogram, peaks that correspond to microcalcifications have a higher overall steepness than normal tissue structures. Moreover, the peaks of some film artifacts are typically steeper than microcalcification peaks. The steepness parameter is obtained by using the first derivative of the area sequence in a manner that results in higher values for steeper peaks.

4) Distinctness. In many cases, the normal breast tissue in a mammogram has a grainy appearance due to a large number of contiguous normal microstructures. Although microcalcifications may be clustered occasionally in close proximity to each other, they are more distinct and separate from each other as well as from normal

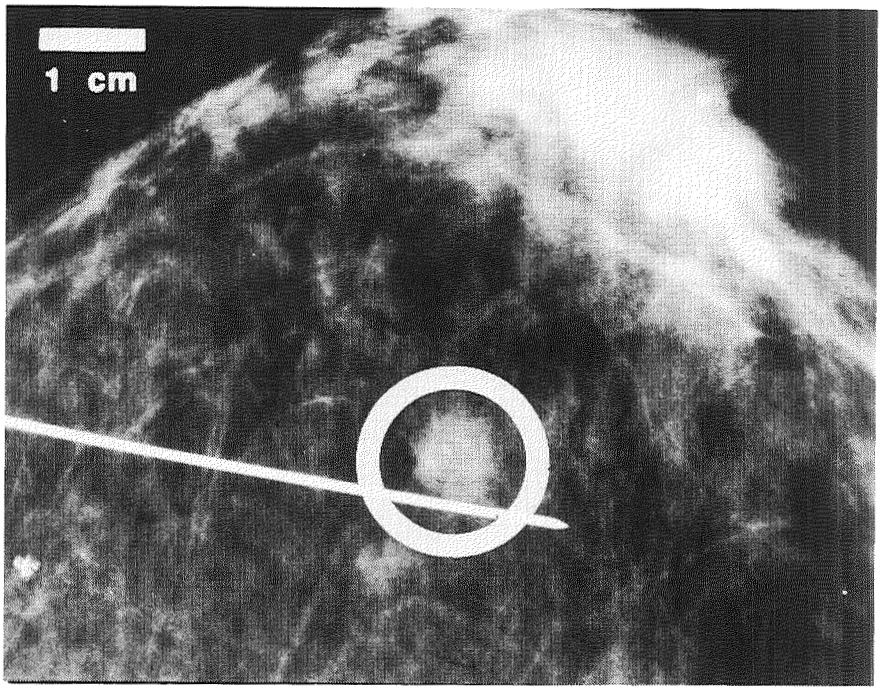

Figure 1. Photographic enlargement of a mammogram analyzed in this study. A microcalcification cluster (circled) is shown.

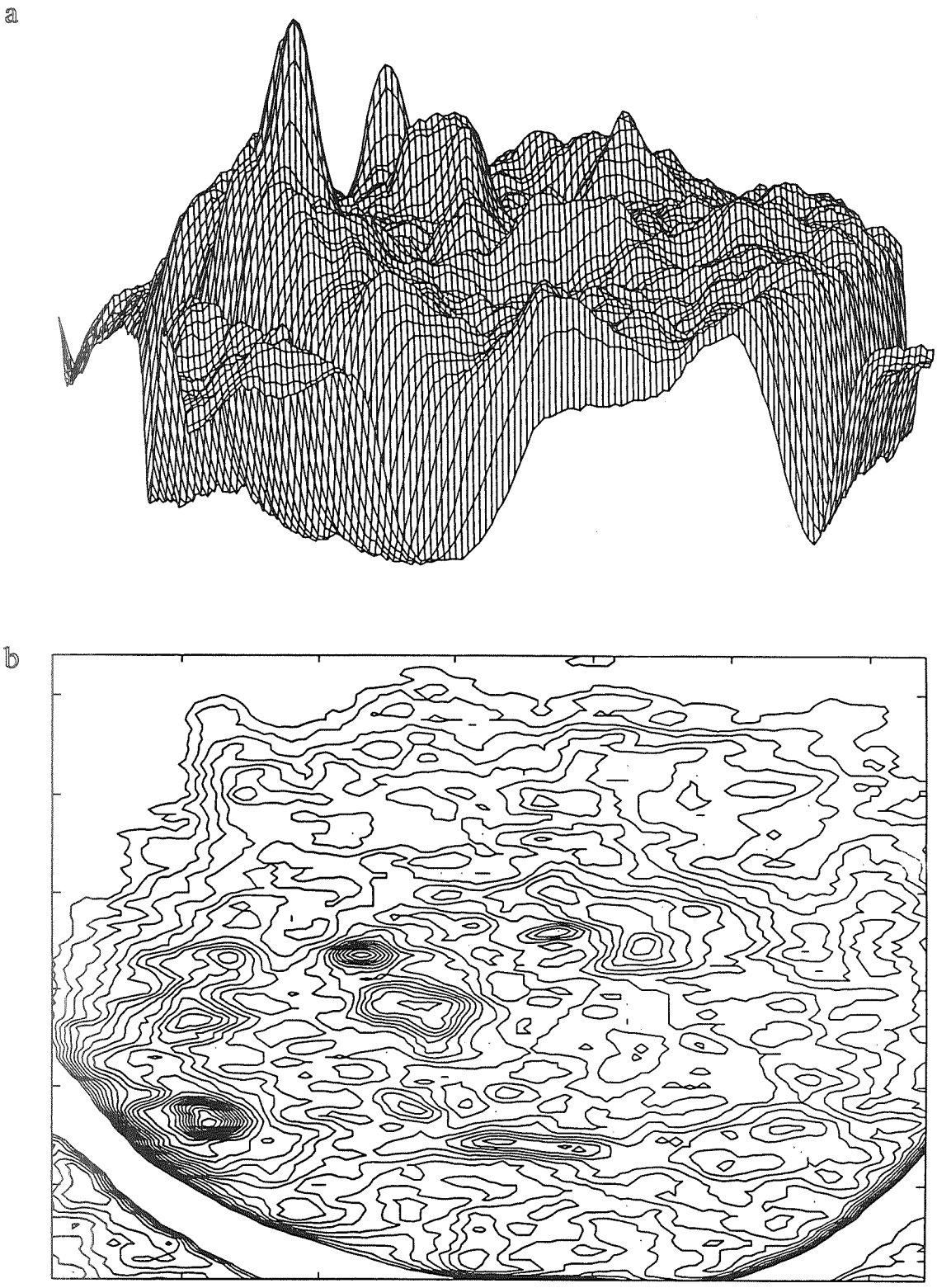

Figure 2. (a) Intensity surface plot of the region containing a microcalcification cluster shown in Fig. 1; (b) Iso-intensity contour plot derived from the same region.

microstructures. The distinctness of a peak is set to the number of contour levels between the tip and the level where its contour merges with that of the nearest peak.

5) Compactness. The edge morphology of a microcalcification is a significant visual cue. For each peak, this morphological information is obtained by using the characteristic contour of a peak obtained just above the merging level. The compactness feature is computed using the ratio of the perimeter to the area of the characteristic contour. Compactness is a standard morphological descriptor that has a value of 1 for a circle and increases as the shape becomes more irregular. The compactness of some types of artifacts and most normal tissue structures is relatively higher than that of microcalcifications.

Each peak is characterized with the 5 features that the algorithm extracts from the raw mammogram data and the discrimination between microcalcifications and other structures is based on these features. The discrimination can be performed with conventional Bayesian classification, standard feedforward neural networks [16], or specialized neural networks [e.g. 17]. In this study the Bayesian classifier was used and adequate results were obtained.

The decision parameters of the classifier were determined on 3 mammograms that formed the training set. The digitized training mammograms formed a data-base of 64 image segments containing more than 1000 microstructures that had a size of interest (less than 2 mm wide). The training set contained 5 of the microcalcification clusters indicated by the radiologist. The distributions of the features were obtained for the populations of microstructures within the indicated clusters (detection class) and for the population of peaks in the rest of the mammogram (rejection class) separately. The decision thresholds were set in order to maximize the discrimination between the detection class and the rejection class. A cluster was indicated by the algorithm when 3 or more microcalcifications occurred in an area of less than $1cm^2$ using the 5 features and a two-phase data reduction approach.

EVALUATION

The performance of the algorithm was evaluated on the 6 other mammograms that formed the test set. The digitized test set resulted in 84 image segments containing more than 1200 candidate microstructures. The test set contained 8 of the clusters indicated by the radiologist and the algorithm detected all 8. In addition, the algorithm detected 1 false positive cluster in one mammogram. Therefore, on the test set the sensitivity was 100% with 0.17 false clusters per mammogram.

In pattern recognition applications, the performance of an approach is measured by the balance of false negatives and false positives that it can provide. Almost any algorithm can be made sensitive enough to detect all events of interest (no false negatives). However, increasing the sensitivity generally reduces the specificity and causes a larger number of false positives. Therefore, in many pattern recognition applications the false positive rate associated with a desirable sensitivity level is used as a measure of performance. A sensitivity level of about 95% or more is desirable in early breast cancer detection. For such a high sensitivity level, the 0.17 false clusters per mammogram obtained with this algorithm provide a considerably better specificity than 5 or more false positive clusters per mammogram obtained with other algorithms.

This algorithm was developed specifically to detect microcalcifications based on the radiographic visual evaluation criteria. These criteria were computationally expressed as features extracted from the raw data without using enhancement. The use of the contour plot provided a convenient technique for computing the features without using preset arbitrary analysis kernels. In this manner, all measurements were obtained using natural morphological contours of microcalcification peaks. The decision thresholds were applied to features and not to the intensity data. Appropriate values of these thresholds were determined using a large number of diverse microstructures. Therefore these thresholds did not depend on local statistics, they held across mammograms and did not have to be adjusted for each mammogram separately. Once the thresholds were set using a representative training set, the algorithm operated in a fully automated manner without human supervision. The algorithm will be further validated on more than hundred mammograms during use in the Department of Radiology of The Johns Hopkins Hospital.

CLINICAL IMPLEMENTATION

The automated system based on this algorithm will be a reliable diagnostic tool that can assist radiologists in early breast cancer detection on mammograms. The system will be made of a scanner, a low-cost workstation, a high-resolution display and a printer for hard copies of results. The speed of the workstation will allow one mammogram to be analyzed in less than 5 minutes. The fully automated operation of the algorithm ensures that the system will not introduce an additional burden to radiologists.

The automated system will be able to analyze a mammogram without human supervision; however, it will be designed to benefit from the experience of radiologists across time. This will be possible with training software that will be available to radiologists or technicians. Occasionally, a human operator will indicate to the system the location of false negative or false positive microcalcifications, using the keyboard or mouse. The training software will automatically adjust the operational parameters of the system to detect the microcalcifications that were missed and to reject the false positive structures. This will be achieved with minimal change on past correct performance because the adjustment will be made by taking into account not only the currently indicated structures but an archive of some previously encountered structures. This archive will contain the features of a large number of microcalcifications as well as other structures (normal tissue, artifacts, etc.) that were located very close to the decision boundary between these two classes. Therefore, the structures in the archive will be those that would be affected first by changes in operational parameters. The training software will automatically optimize the discrimination based on the past examples and the currently indicated structures. In this adjustment, the weight given to current structures will be user-selectable.

The automated operation of this system is especially suited for clinics that have to screen a large number of mammograms every day. In such a clinical setting, the system can operate virtually all the time, as long as a technician is available for feeding the mammograms to the scanner. An automated feeding instrument can also be conceived. Assuming 10 hours of operation per day and a worst case of 5 minutes per mammogram, about 120 mammograms a day can be screened by the system without requiring any time from the radiologist. When the results of the automated system are available, the expert radiologist will focus on the regions where the system indicated microcalcification clusters in each mammogram to confirm the results. For the purpose of quality control, the radiologist might also screen some regions that were cleared by the system on several mammograms. The expected clinical benefits are: i) accurate detection of subtle signs of breast cancer that might be missed by radiologists and, ii) significant reduction in the amount of time that radiologists spend for screening mammograms. Currently, due to the subtlety of early breast cancer signs, radiologists use a magnifying glass to screen mammograms. The time required for the visual interpretation of a complete mammogram can often reach 15 minutes and in some cases it can take up to 30 minutes. The automated system is expected to reduce the time required of the radiologist by an order of magnitude.

REFERENCES

1. American Cancer Society, Ca Cancer J. Clin. 33, 226 (1982).
2. D.B. Kopans, J.E. Mayer and N. Sadowsky, N. Engl. J. Med. 310, 960 (1984).
3. L. Bassett, "Mammographic analysis of calcifications", Radiologic Clin. N. Amer., 30(1), 95 (1992).
4. P.C. Stomper, J.C. Connolly, J.E. Meyer et al, "Clinically occult ductal carcinoma in situ detected with mammography: Analysis of 100 cases with radiologic-pathologic correlation", Radiology, 172, 235 (1989).
5. E.A. Sickles, "Mammographic Features of 300 consecutive nonpalpable breast cancers", Am. J. Roetgenol., 146, 661-663 (1986).
6. M.J. Smith, "Error and variation in diagnostic radiology", pub. Charles C. Thomas, Springfield, IL (1967).
7. L. Kalisher, "Factors influencing false negative rates in xeromammography", Radiology 133, 297-301 (1979).
8. J.E. Martin, M. Moskowitz, and J.R. Milbrath, "Breast cancer missed by mammography", Amer. J. Roentgen. 132, 913-918 (1982).
9. H.P. Chan, K. Doi, C.J. Vyborny, K.L. Lam and R.A. Schmidt, "Computer-aided detection of microcalcifications in mammograms", Investigative Radiology 9, 664 (1988).
10. B.W. Fam, S.L. Olson, P.F. Winter, F.J. Scholz, "Algorithm for the detection of fine clustered microcalcifications on mammograms", Radiology 169, 333-337 (1988).
11. D.H. Davies and D.R. Dance, "Automated computer detection of clustered microcalcifications in digital mammograms", Physics in Medicine and Biology 35(8), 1111-1118 (1990).

12. E.A. Patrick, M. Moskowitz, V.T. Manshukhani and E.I. Gruenstein, "Expert learning system network for diagnosis of breast calcifications", Investigative Radiology 26, 534-539 (1991).
13. R.C. Gonzales and P. Wintz, "Digital Image Processing", 229-232, pub. Addison-Wesley, Reading, MA (1987).
14. I.N. Bankman, V.G. Sigillito, R.A. Wise, and P.L. Smith, "Feature-based detection of the K-complex wave in the human electroencephalogram using neural networks", IEEE Trans. on Biomed. Eng. (1992), in press.
15. B.H. Jansen, "Artificial neural nets for K-complex detection", IEEE Eng. Med. Biol., 9, 50-52, (1990).
16. D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, "Learning internal representations by error propagation", MIT press, Cambridge, MA, 318-328, (1988).
17. I.N. Bankman, J. Sadowsky, V.G. Sigillito, "A Neural Network for Adaptive Surfaces and Closed Decision Boundaries", Proc. 25th Annual Conf. on Information Sciences and Systems, 25, 848-853, (1991).