Open access • Journal Article • DOI:10.1007/BF02742937

Computer-assisted reading of mammograms. — Source link

Nico Karssemeijer, Jan H. C. L. Hendriks

Published on: 01 Jan 1997 - European Radiology (Eur Radiol)

Topics: Digital mammography, Mammography and Breast cancer screening

Eur. Radiol. 7, 743–748 (1997)

Original article

Computer-assisted reading of mammograms

N. Karssemeijer, J. H. C. L. Hendriks

Department of Radiology, University Hospital Nijmegen, P.O. Box 9101, Nijmegen, NL-6500 HB, The Netherlands

Received 3 June 1996; Revision received 4 September 1996; Accepted 11 September 1996

Abstract. Techniques developed in computer vision and automated pattern recognition can be applied to assist radiologists in reading mammograms. With the introduction of direct digital mammography this will become a feasible approach. A radiologist in breast cancer screening can use findings of the computer as a second opinion, or as a pointer to suspicious regions. This may increase the sensitivity and specificity of screening programs, and it may avoid the need for double reading. In this paper methods which have been developed for automated detection of mammographic abnormalities are reviewed. Programs for detecting microcalcification clusters and stellate lesions have reached a level of performance which makes application in practice viable. Current programs for recognition of masses and asymmetry perform less well. Large-scale studies still have to demonstrate if radiologists in a screening situation can deal with the relatively large number of false positives which are marked by computer programs, where the number of normal cases is much higher than in observer experiments conducted thus far.

Key words: Breast – Mammography – Diagnosis – Digitization

Introduction

It is estimated that in current breast cancer screening programs radiologists do not detect approximately 25 % of the cancers which are visible on retrospective review [1–6]. Moreover, if minimal signs identified on previous screening mammograms are also taken into account, estimates of the number of cancers not reported in screening even range up to 50 % depending on the subjective criteria used by the radiologists performing retrospective reading. The problem of screening errors should be viewed in light of the high specificity which is required in screening. For instance, in the Dutch screening program only approximately 4–5 of 1000 women in the screened population have breast cancer, whereas the positive predictive value of screening is currently approximately 55 %. Given the fact that mammographic signs indicating early stages of breast cancers are often subtle, and considering the speed at which screening radiologists usually read their cases, this may be regarded as a remarkable achievement. It is noted that this high specificity is only achieved by having all cases read by two radiologists.

Part of the problem of missed lesions is likely to be due to inadequate search [7]. If a subtle abnormality, such as a small cluster of microcalcifications, is not hit by foveal vision, it may easily be overlooked. However, visual search patterns recorded from radiologists during reading of chest X-rays and mammograms revealed that many missed abnormalities are actually fixated longer than normal areas [8]. This suggests that another perceptual mechanism plays an important role. Apart from that, more than half of the radiological errors in mammography screening are found to be interpretation errors, i.e. malign lesions which were consciously judged by a radiologist and reported benign.

Digitization of the mammographic imaging procedure allows the use of computers to aid radiologists in reading mammograms. An expanding number of research groups is active in this field. Most of this work is aimed at developing methods for detection of abnormalities such as microcalcification clusters, densities and stellate lesions. These methods can be applied to mark suspicious areas in mammograms when they are read, in order to reduce detection errors. This approach has turned out to be successful in a number of studies [9–11]. However, neither of these studies has yet demonstrated the impact in a real screening program, where the fraction of abnormal cases is only a few percent. Other approaches in computer-aided diagnosis (CAD) in mammography aim at helping radiologists in inter-

Correspondence to: N. Karssemeijer

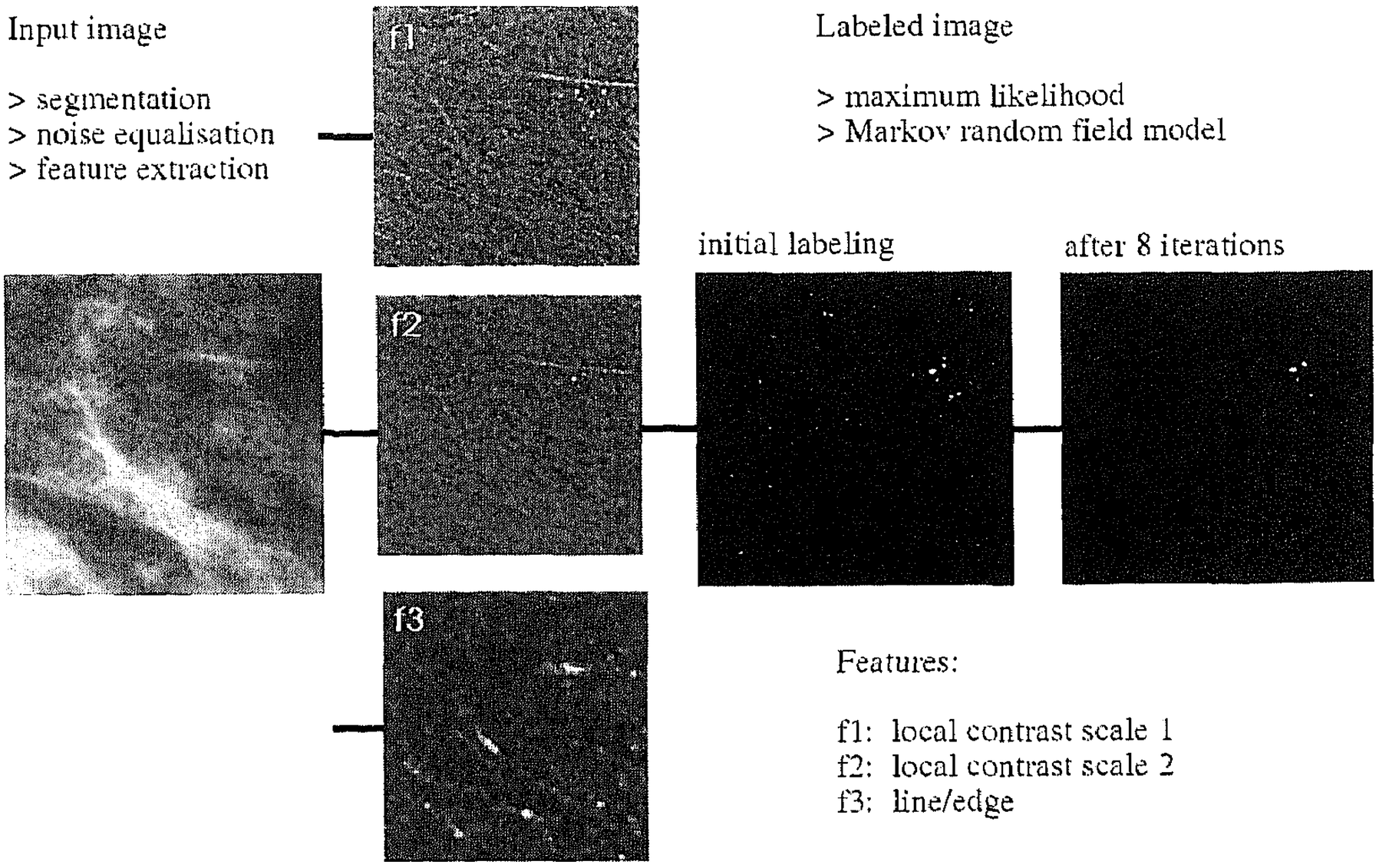

Fig. 1. Overview of a method for automated detection of microcalcification clusters. After noise equalization, three feature images are computed, which are combined in an initial labeling of microcalcification *(white)* and background, line or emulsion artefact pixels *(black)*. Iteratively, the labeling is optimized by removing unlikely candidate pixels, using a model which allows to maintain an increased sensitivity within clusters

preting lesions. Detection is a lesser problem here, because the position of abnormalities can be constrained by annotations of the radiologist. Most of the research in classification methods is directed at characterization of microcalcification clusters. Pattern recognition methods, such as neural networks, seem to be very well suited for this problem, which requires combination of evidence from many sources. Detection of microcalcifications remains an important issue, because computerized classification requires detection and segmentation first.

The aim of this paper is to describe techniques developed for detection of abnormal mammographic patterns, where the focus is on clustered microcalcifications and stellate lesions. Furthermore, it is discussed how the performance of such techniques can be objectively assessed.

Methods for detection

Digital mammograms can be obtained by digitization of conventionally recorded film-screen systems or by using digital acquisition devices. It is good to realize that automated pattern recognition programs are often not very robust when images come from different sources. Each device has its own characteristics with respect to noise and contrast transfer, and the positioning of labels and markers on mammograms may differ. This may cause unexpected problems for computer programs, which are often tuned to one particular image data set. To minimize such problems it is important to develop methods which are invariant for resolution and grey-scale conversions, and which use reliable methods for segmentation of breast tissue from the background. The latter step is common to almost any program used in breast image processing.

Methods for detection of abnormalities in mammograms can roughly be classified into sampling- and region-based approaches. In the first approach, local image features are calculated at a set of regularly spaced points across the whole breast area. Then, for each point a measure of suspiciousness is computed from these features and points are subsequently grouped into regions marked as normal or suspicious. In the region-based approach, the initial step of the program is creation of a subdivision of the breast image into regions. Then, for each of these regions image features are calculated and combined into a measure for the degree of suspiciousness.

Detection of microcalcifications

Recognition of microcalcification clusters has been studied by many researchers [12–19]. All methods have in common that one or more filters are used to determine local contrast at each pixel inside a region of interest, usually representing the whole breast. Microcalcifications have high local contrast. However, also other high-contrast structures exist such as vessel walls and thin strings of connective tissue. Furthermore, peaks in the image noise may be hard to distinguish from microcalcifications. Simply selecting pixels with high local contrast is not an appropriate way, therefore, to detect microcalcifications, because this would yield far too many false positives. Automated detection methods differ in the way they deal with this problem. Often, many candidate spots are selected on the basis of local contrast alone, and a second processing stage removes false positives using models which represent properties of noise and normal tissue.

The method which was developed at our institute is outlined in Fig. 1. In a preprocessing stage the image

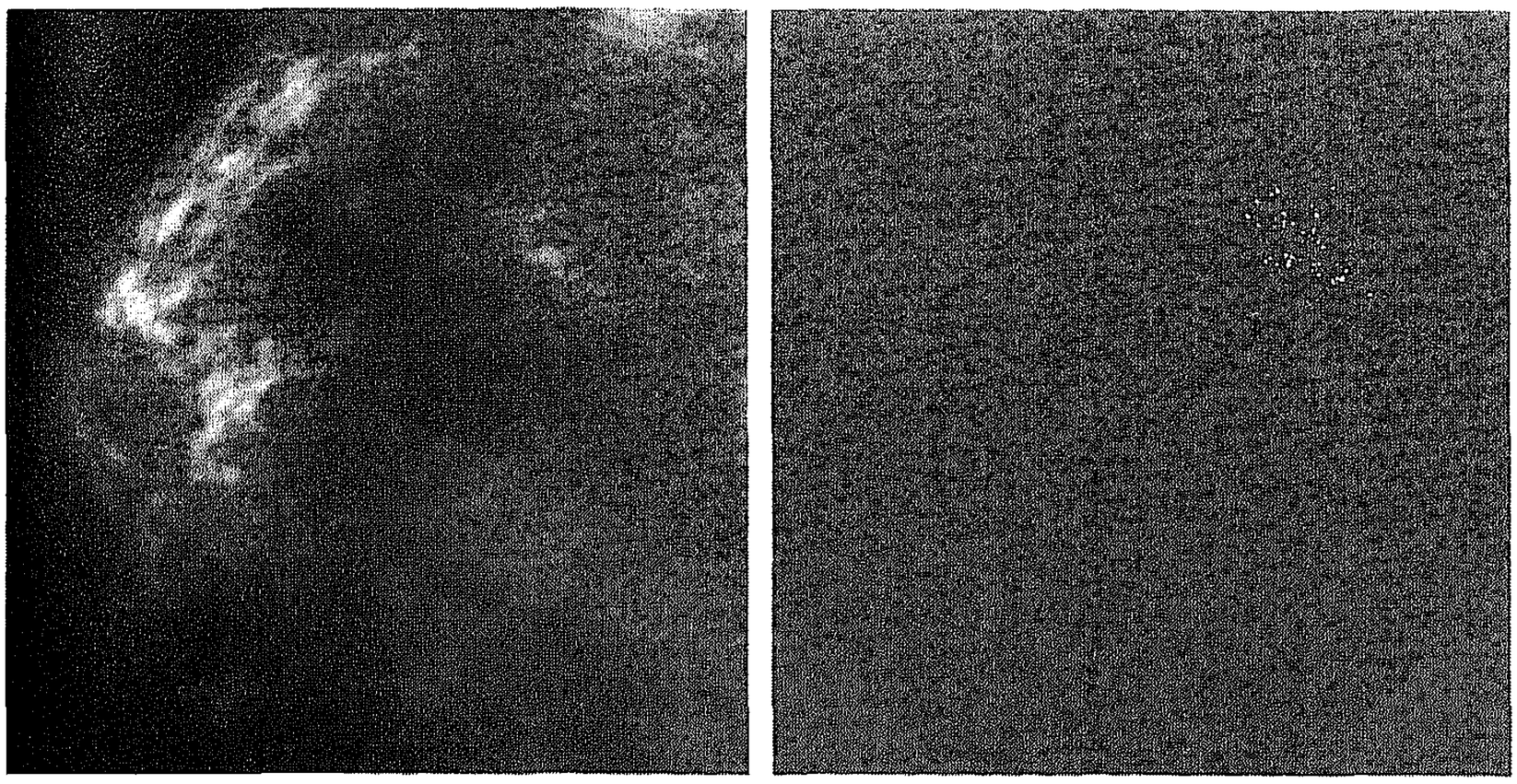

Fig. 2. An example of a cluster of microcalcifications which is automatically detected

noise is estimated as a function of signal intensity, which allows conversion of the mammograms to a grey scale with known constant noise level. This noise equalization procedure greatly facilitates further analysis [16], because it makes the detection algorithm less dependent on image acquisition. The processed image is subsequently used as input to three feature detectors, which compute local contrast at two spatial scales and extract a measure which indicates if a linear line or edge structure is likely to be present. These three features are combined in a statistical way to label all pixels in the image as background, line or film emulsion artefact (black) or as microcalcification (white). This initial labeling is performed at a very sensitive level in order not to miss any true microcalcifications. Finally, each pixel is relabeled in an iterative scheme, making use of the information available in its neighbourhood. Only pixels with a high likelihood of being part of a true microcalcification survive in this process, where the likelihood depends on how many other microcalcifications there are in their neighbourhood. An example of a mammogram in which a cluster of microcalcifications was detected is shown in Fig. 2.

Detection of masses

In the majority of cases missed by screening a mass is involved [4], which is the most important mammographic sign for detection of invasive breast cancer. Masses may be hard to detect as they are easily obscured by the fibroglandular tissue, especially when the tumor is still small. Computerized detection of masses in mammograms has been investigated by a number of groups. Lau and Bischof [20] and Ng and Bischof [21] try to determine the presence of masses by comparing brightness and texture features in corresponding regions of the left and right mammograms. Registration is performed by mapping three control points on the breast boundary, the nipple and the two chest wall points, and by calculating the displacement of all other points by some interpolation rule. Giger et al. [25] and Nishikawa et al. [24] designed a method to detect masses based on subtraction of left and right mammogram after automatic alignment. Potentially suspicious regions are identified in the subtraction images and classification of these regions is performed on the basis of image features determined in these regions. Results range from a true-positive detection fraction of 90 % for lesions larger 2 cm to 30 % for tumors of approximately 1 cm, both at a false-positive level of 2 FP/image. Miller and Astley [22] also investigated asymmetry detection. They recognized the problem of inadequate registration generating false asymmetries, and analysed methods which radiologist use in order to learn from their experience. They found that many of the radiologist's comparisons have a regional basis, considering four breast quadrants and the glandular region. As the majority of the cancers are located in the glandular region, where they may be obscured, radiologists have special consideration for the shape of the glandular disc. The researchers attempt to segment fat from non-fat regions and to detect masses by determining left-right asymmetry in the shapes of the non-fat regions. Brzacovic et al. [23] describe a multiscale mass detection scheme based on fuzzy logic. On a test set of 12 images showing an irregular mass their method correctly labeled the mass in 8 cases, but no false-positive rate was reported. Finally, in a recent study Petrick et al. [26] suggest the use of a new contrast enhancement filter followed by edge detection to generate target signals, which are then classified by a statistical or neural network classifier.

Detection of stellate lesions

It appears that most malignant densities have irregular shapes and that they are frequently surrounded by a radiating pattern of linear spicules. Sometimes, the central density is faint or absent. In those cases the stellate pattern of spicules is the most important sign. Kegelmeyer et al. [11] and Kegelmeyer [27] describe a method for detection of stellate lesions. The method is based on the analysis of histograms of local edge orientations and application of the Law's texture measures. Features are combined using a binary decision tree to label pixels

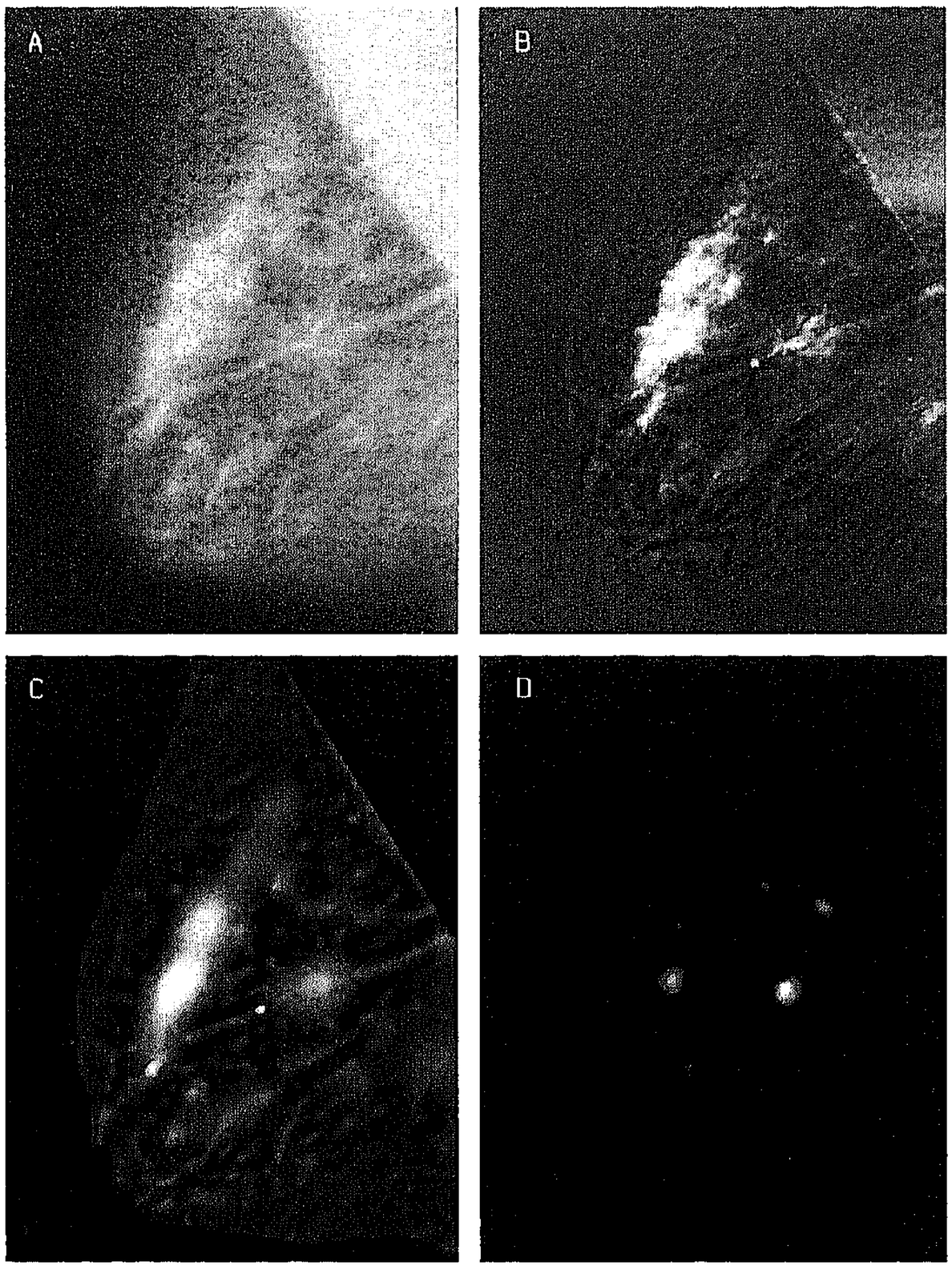

Fig. 3 A–D. A stellate lesion automatically detected. The pectoral muscle of the mammogram shown in **A** is recognized and removed, and intensities near the skin line are compensated for by reduced breast thickness (**B**). **C** Bright areas detected at different scales determine the size of regions to be analysed for spiculation. **D** The result shows the measure of suspiciousness which was computed at each image site. The bright area marks the position of a malignant stellate lesion

as normal or abnormal. This pixel-based approach for recognition of stellate lesions seems to be superior to the region-based approach described by Woods and Bowyer [28]. Pixelwise classification based on local features also forms the basis of a method for detection of stellate distortion developed at our institute [31]. In this method detection is performed by statistical analysis of a map of line-based pixel orientations. The idea is that if a strong increase in pixels pointing to a given regions is found, this region is suspicious, especially if such an increase is found in many directions. No attempt is being made to explicitly identify spicules. The calculations are performed at a grid of sites inside the imaged breast area. Afterwards, neighbouring sites with a high level of suspiciousness are linked up to form regions. Before calculation of the pixel orientation map, the breast tissue is segmented from the background, and in the oblique views the pectoral muscle is marked. An example of this segmentation, which is performed fully automatically, is shown in Fig. 3 b. It used to perform a filtering step in which the pectoral muscle is removed and in which the decrease of image intensities near the breast skin line is compensated for. The processed image is used as input to a bank of filters for detection of bright areas at a range of spatial scales (Fig. 3 c). When such a bright area is present, its scale is used to set the size of the surroundings to be analysed for detection of spiculation. In case a small density is present, the program analyses a smaller region that in the case of a larger density.

Pixel orientations are estimated using a new method based on Gaussian scale-space theory. If a line-like structure is present at a given pixel, this method provides an accurate estimate of its orientation, whereas in other cases the image noise generates some random output. The orientation estimates are used to construct two operators which respond to radial patterns of straight lines. The first one is defined to measure the total number of pixels with directions pointing to a test area. If this number is significantly more than would be expected on the basis of randomness, the area may be suspicious. However, if an increase in the number of pixels oriented towards a region is found in a few directions only, it is not very likely that the site being evaluated belongs to the centre of a stellate pattern. On the other hand, if evidence for spicules is found in many directions, this should increase the likelihood of a stellate structure being present. To represent this property, a second operator is constructed, measuring the uniformity of the orientation map. The two features defined are combined to form an image representing the degree of suspiciousness at each pixel. Figure 3 d shows an example where bright spot is at the location of a histologically verified malign stellate distortion.

Performance measurement

With the strong increase in research in computer-aided diagnosis, objective measurement of the performance of computer programs which detect lesions has become an important issue. There are two main problems. The first one is that most researchers use their own image databases, which makes it hard to compare published results. To overcome this problem, common databases with digital mammograms are being distributed now, or will become available soon [16, 29]. The second problem is standardization of performance measurement, given that reliable annotations are part of the database.

A method which is widely used in radiology for measuring detection performance is receiver-operating-characteristics (ROC) analysis [30]. In this method an image set is collected which contains both normal and abnormal cases. Then, each image is rated in an observer performance study according to the confidence of a reader that the image is abnormal. Experimental results are plotted as an ROC curve, representing the true-positive fraction as a function of the false-positive rate. The ROC analysis has been applied to evaluate the performance of radiologists with and without computer-assisted reading.

A problem with ROC analysis is that it cannot deal with multiple lesions per case and that the locations of

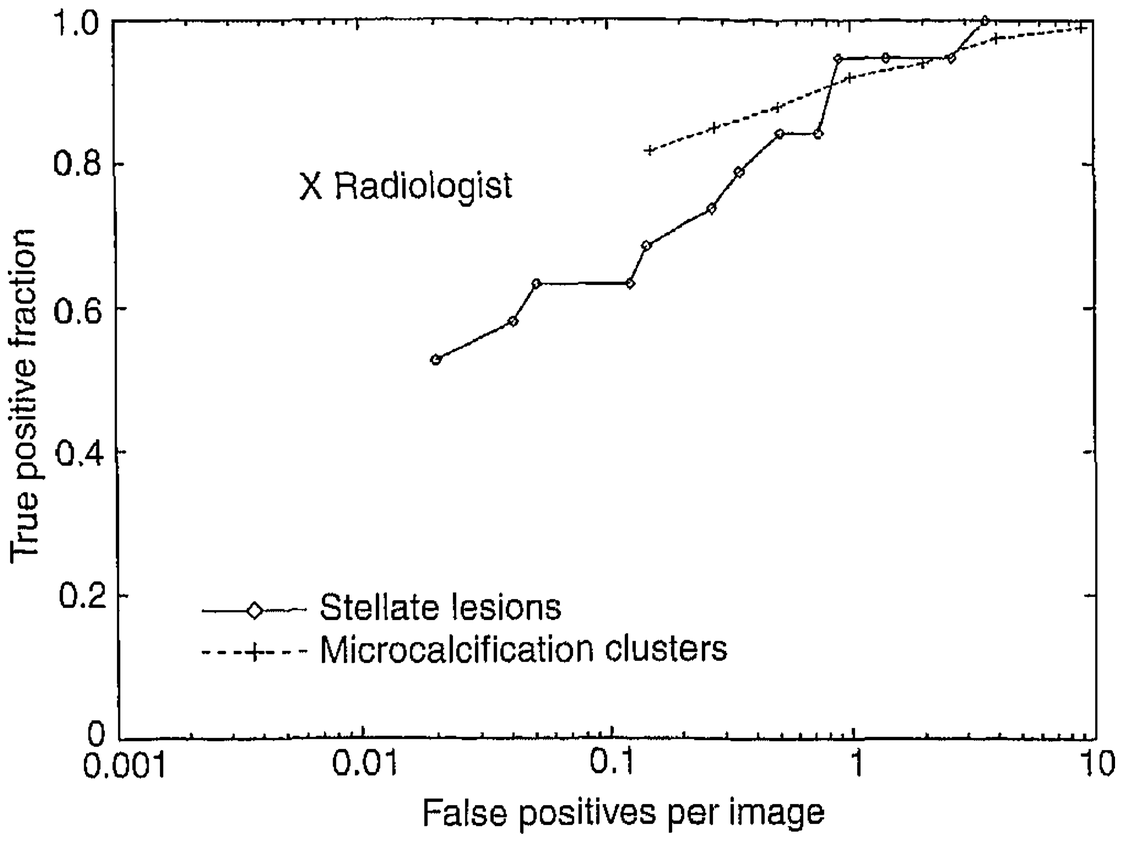

Fig. 4. Free-response receiver-operating-characteristics curves showing results for automated-detection microcalcification clusters and stellate lesions. Also, a rough estimate of the average performance of a radiologist in breast cancer screening is marked (X)

lesions reported by a reader or program are not verified. For analysis of detection performance in complex images, such as mammograms, this is a severe drawback. In case of microcalcifications, for instance, multiple clusters often occur. Moreover, at the current stage of development, computerized detection of mammographic lesions often results in more than one area signaled per image. For this reason free-response ROC (FROC) is a more appropriate tool for evaluating the performance of automated detection [32]. In this approach the fraction of true positives is calculated as a function of the number of false positives per image. To generate such a curve, a parameter can be adjusted in most algorithms to trade off sensitivity against specificity. A disadvantage of the use of FROC curves is that procedures for performing statistical analysis are still under development.

When judging experimental results, it is important to realize that definitions used for true- and false-positive detections often differ. To check whether a suspicious area marked by a program corresponds with an annotated lesion, often some overlap criterion is used. If the overlap with the annotation is large enough, the detected area is counted as a hit; otherwise, it is a false positive. Another criterion which can be applied to determine the hit rate uses the most suspicious point in a marked region. A true positive is only counted if this point lies within the annotated region. In general, we found that different criteria give similar results when the false-alarm rate is low. On the other hand, results may differ significantly when the number of false positives per image gets larger than three.

Figure 4 shows FROC curves for detection of microcalcifications and stellate lesions. Both curves were determined on common datasets with software developed at our institute. The method for stellate lesion detection was applied to 50 mammograms taken from the MIAS database [29], which are all cases from the UK breast cancer screening program. The images were resampled from the original 50 μm resolution to 200 μm/pixel before processing. At a rate of one false positive per image, a detection sensitivity of more than 90% is achieved. The method for detection of microcalcifications was tested on 40 digital mammograms digitized in our institute with a 12-bit CCD camera at 100 μm/pixel. At one false positive per image 95% of the clusters were detected in this database. The 40 images were made available for other research groups and have been used in a number of published studies [17–19]. Microcalcification detection results that we obtained on this database had not yet been improved by others.

Conclusion

The performance of pattern recognition programs which have been developed to aid radiologists in detecting breast cancer is reaching a level now that application seems to be becoming worthwhile. The sensitivity of methods for detecting, microcalcification clusters and stellate lesions is relatively high at a false-positive rate which may be acceptable. For densities and asymmetry the performance of computerized detection is still low. For instance, at a specificity of 1 FP/image a true-positive fraction of only 30% is reported by Nishikawa et al. [24] for masses with diameters between 8 and 14 mm. Similar results were obtained by Petrick et al. [26] for masses of unknown size. For stellate lesions and microcalcification clusters the sensitivity is approximately 90% at the same specificity. No results have been published yet which come close to the performance of human readers. For example, assuming a positive predictive value of 20% and an incidence of 0.5% [33], it can be calculated very roughly that a radiologist in breast cancer screening reports approximately 20 false positives per 1000 cases. This corresponds to 0.01 FP/image, given that there are two oblique views taken per case. With an estimated sensitivity of approximately 75% this point can be marked in the FROC plot (Fig. 4). This gives an impression of the different levels of performance of radiologists and computer programs. Of course, the performance level of radiologists also includes detection of masses, and with respect to microcalcifications it should be noted that radiologists do not only detect them, but also distinguish benign from malign types. The latter is not reflected in the FROC curve of the computer readout.

Although there is convincing evidence now that prompting of suspicious areas found by computer programs can increase in the level of performance of radiologists significantly, it is not known yet how radiologists in a real screening situation can deal with large numbers of false positives. Observer performance studies which have been conducted up to now use datasets in which the fraction of abnormal cases is far larger than in screening. Large trials need to be performed to demonstrate that computer-assisted reading is beneficial.

References

1. Baines CJ, McFarlane DV, Miller AB (1990) The role of the reference radiologist: estimates of inter-observer agreement and potential delay in cancer detection in the national breast screening study. Invest Radiol 25: 971–976
2. Bird RE, Wallace TW, Yankaskas BC (1992) Analysis of cancers missed at screening mammography. Radiology 184: 613–617
3. Harvey JE, Fajardo LL, Inis CA (1993) Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blinded interpretation. AJR 161: 1167–1172
4. Dijck JAM van, Verbeek LM, Hendriks JHCL, Holland R (1993) The current detectability of breast cancer in a mammographic screening program. Cancer 72 (6): 1933–1938
5. Gale A, Wilson ARM, Roebuck EJ (1993) Mammographic screening: radiologic performance as a precursor to image processing. In: Proc First int Workshop on Digital Mammography, SPIE vol 1905, pp 458–464
6. Savage CJ, Gale AG, Pawley EF, Wilson ARM (1994) To err is human, to compute divine? In: Proc Second Int Workshop on Digital Mammography, Elsevier Int Congr Series 1069, pp 405–414
7. Vyborny CJ (1994) Can computers help radiologists read mammograms? Radiology 191: 315–317
8. Krupinski EA, Nodine CF (1994) Gaze duration predicts the locations of missed lesions in mammography. In: Gale AG (ed) Digital mammography. Elsevier Int Congr Series 1069, pp 399–405
9. Chan HP, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam KL, Ogura T, Wu Y, Macmahon H (1990) Improvement in radiologist's detection of clustered microcalcifications on mammograms. Invest Radiol 25: 1102–1110
10. Astley S, Hutt I, Adamson S, Rose P, Miller P, Boggis C, Taylor C, Valentine T, Davies J (1993) Automation in mammography: computer vision and human perception. In: Proc First Int Workshop on Digital Mammography, SPIE 1905, pp 716–730
11. Kegelmeyer WP Jr, Pruneda JM, Bourland PD, Hillis A, Riggs MW, Nipper ML (1994) Computer-aided mammographic screening for spiculated lesions. Radiology 191: 331–337
12. Chan HP, Doi K, Vyborny CJ, Lam KL, Schmidt RA (1988) Computer-aided detection of microcalcifications in mammograms. Invest Radiol 23: 664–671
13. Davies DH, Dance DR (1990) Automatic computer detection of clustered calcifications in digital mammograms. Phys Med Biol 35: 1111–1118
14. Astley SM, Taylor CJ (1990) Combining cues for mammographic abnormalities. Proc First British Mach. Vision Confl, pp 253–258
15. Karssemeijer N (1992) A stochastic model for automated detection of calcifications in digital mammograms. Image Vision Comput 10: 369–375
16. Karssemeijer N (1993) Adaptive noise equalization and detection of microcalcification clusters in mammography. IJPRAI 7: 1357–1367
17. Kegelmeyer WP Jr, Allmen MC (1994) Dense feature maps for detection of calcifications. In: Gale AG (ed) Digital mammography. Elsevier Int Congr Series 1069, pp 3–12
18. Netsch T (1996) Detection of microcalcification clusters in digital mammograms: a scale-space approach. In: Arnolds B et al. (eds) Digitale Bildverarbeitung in der Medizin. GMDS, Freiburg
19. Strickland RN, Hahn HI (1996) Wavelet transforms for detecting microcalcifications in mammograms. IEEE Trans Med Imaging 15: 218
20. Lau TK, Bischof WF (1991) Automated detection of breast tumors using the asymmetry approach. Comput Biomed Res 24: 273–295
21. Ng SL, Bischof WF (1992) Automated detection and classification of breast tumors. Comput Biol Res 25: pp 218–237
22. Miller P, Astley S (1993) Detection of asymmetry using anatomical features. Proc First Int Workshop on Digital Mammography, SPIE 1905, pp 433–442
23. Brzakovic D, Brzakovic P, Neskovic M (1993) An approach to automated screening of mammograms. Proc First Int Workshop on Digital Mammography, SPIE 1905, pp 690–701
24. Nishikawa RM, Giger ML, Doi K, Vyborny CJ, Schmidt RA (1994) Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms. In: Bowyer K, Astley S (eds) State of art in digital mammographic image analysis. World Scientific, Machine Perception and AI 9, Singapore, pp 82–103
25. Giger ML, Lu P, Huo Z, Bick U, Vyborny CJ, Schmidt RA, Zhang W, Metz CE, Wolverton D, Nishikawa RM, Zouras W, Doi K (1994) CAD in digital mammography: computerized detection and classification of masses. In: Gale AG (ed) Digital mammography. Elsevier Int Congr Series 1069, pp 281–288
26. Petrick N, Chan HP, Sahiner B, Wei D (1996) An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. IEEE Trans Med Imaging 15: 59–67
27. Kegelmeyer WP Jr (1992) Computer detection of stellate lesions in mammograms. Proc Biomedical Image Processing and 3-D Microscopy, SPIE 1660, pp 446–454
28. Woods KS, Bowyer KW (994) Computer detection of stellate lesions. In: Gale AG (ed) Digital mammography. Elsevier Int Congr Series 1069, pp 221–230
29. Suckling J, Parker J, Dance DR, Astley S, Hutt I, Boggis CRM, Ricketts I, Stamatakis E, Cerneaz N, Kok SL, Taylor P, Betal D, Savage J (1994) The mammographic image analysis society digital mammogram database. In: Proc Second Int Workshop on Digital Mammography. Elsevier Int Congr Series 1069, pp 375–378
30. Swets J, Pickett RM (1982) The evaluation of diagnostic systems: methods from signal detection theory. Academic Press, New York
31. Karssemeijer N (1995) Detection of stellate distortion in mammograms using scale space operators. In: Bizais Y, Barillot C, Paola R di (eds) Information processing in medical imaging. Computational imaging and vision 3. Kluwer, Dordrecht, pp 335–346
32. Chakraborty DP, Winter LHL (1990) Free response methodology: alternate analysis and a new observer-performance experiment. Radiology 174: 873–881
33. Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S (1993) Report of the international workshop on screening for breast cancer. JNCI 85: 1644–1656