Computerized Classification Can Reduce Unnecessary Biopsies in BI-RADS Category 4A Lesions

Leichter Isaac[1,2], Lederman Richard[3], Buchbinder Shalom[4], Srour Yossi[1], Bamberger Philippe[1], and Sperber Fanny[5]

[1] Siemens Computer Aided Diagnosis, PO Box 45202, Jerusalem 91450, Israel
{isaac.leichter, yossi.srour, philippe.bamberger}@siemens.com
[2] Department of Medical Engineering, Jerusalem College of Technology, JCT PO Box 16031, Jerusalem 91160, Israel
[3] Department of Radiology, Hadassah University Hospital, P.O. Box 54202, Jerusalem 91540, Israel
rlederman@hadassah.org.il
[4] Department of Radiology, Staten Island University Hospital Staten Island, 475 Seaview, Staten Island, New York 10305-3498
sbuchbinder@siuh.edu
[5] Department of Radiology, Tel Aviv Sourasky Medical Center, 6 Weizman St., Tel Aviv 64239, Israel
sperber@tasmc.health.gov.il

Abstract. The objective of the study was to assess the potential of a CAD device with computer aided classification capabilities to reduce interventional procedures for BI-RADS category 4A lesions. 113 such lesions (17 masses, 96 clusters), forwarded for biopsy (103 benign) were analyzed retrospectively by a CAD device that generated descriptors. The device extracted quantitative features characterizing the lesions by shape, margins, size and distribution. Descriptors taken from the BI-RADS lexicon for the appearance of the lesion were generated based on the values of the quantitative features. A paradigm based on the computer generated descriptors was developed to assist in assigning a level of suspicion. The paradigm deemed malignant, all 10 malignant cases of the study (100% sensitivity) and correctly classified 38 of the 103 benign lesions. The CAD-generated descriptors, thus, eliminated 36.9% of unnecessary biopsies without decreasing the sensitivity.

1 Introduction

Computer Aided Diagnosis (CAD) in mammography [1] has received FDA approval and entered the mainstream of clinical practice. Its exact role, however, has yet to be defined [2,3], and its widespread implementation is hindered by the relatively large number of false marks [4]. Also, the current generation of CAD systems serves only as a second reader, designed to avoid missed lesions [5], without offering the radiologist a second opinion regarding the nature of the finding. As with all screening tests, mammography is subject to a lack of specificity, which leads to further evaluation of suspicious findings [6]. The need for breast biopsy, frequently with benign results

Susan M. Astley et al. (Eds.): IWDM 2006, LNCS 4046, pp. 76 – 83, 2006.
© Springer-Verlag Berlin Heidelberg 2006

[7,8], has both a financial and psychological cost, which can be cut by increasing the specificity of diagnosis in mammography.

The addition of classification capabilities could potentially improve the efficacy of such systems by calculating the level of suspicion of any finding either detected by the first tier of the system, or considered suspicious by the radiologist. Several machine learning methods based on neural networks and Support Vector Machines have been applied for the classification of mammographic lesions [9,10]. It was found that for microcalcifications, a classifier based on kernel-based methods, such as Support Vector Machines and Kernel Fisher Discriminant, yielded a significantly better performance than neural network [11].

In this study a classification scheme, based on Kernel Fisher Discriminant, is described, and its use is tested in a subdivision of BI-RADS category 4 cases with both benign and malignant pathologies. BI-RADS category 4 includes findings that do not have the classic appearance of malignancy but have a wide range of probability of malignancy. It is the most problematic and subjective category resulting in a high percentage of benign biopsies. Category 4A is a subdivision, which includes findings with the lowest level of suspicion, for which interventional procedures are nevertheless still recommended. It has been shown that the BI-RADS descriptor categories stratify suspicious micro-calcifications appropriately into intermediate and higher probability of malignancy groups [12]. In this study, an attempt was made to further refine which lesions in this BI-RADS category, in fact, should be sent for biopsy, by the use of computerized descriptors reflecting the appearance of the lesions in the mammogram. The descriptors generated by the CAD device are similar to those used by the BI-RADS lexicon and are familiar to the radiologist.

2 Methods and Material

One hundred and nine cases with 113 lesions (17 masses, 96 clusters) were retrospectively culled from the archives of a university-affiliated facility. All the cases had been prospectively assigned BI-RADS 4A and forwarded for stereo-tactic biopsy. The mean age of the patients was 54.1 ± 8.6 years (range 33–72). The Institutional Review Board at the institution approved the use of these cases for the study, and did not require informed consent because the study was retrospective and patient anonymity was strictly enforced in all aspects of the study. Of the 113 BI-RADS category 4A lesions, 15 masses and 88 clusters proved to be benign at pathology.

The mammograms of the 109 cases were digitized at high resolution (600 dpi, 12 bit) by a prototype CAD device developed by Siemens CAD, Israel [13,14] and the digital images were displayed on the computer screen for further analysis. All 113 lesions were analyzed retrospectively by a radiologist using the CAD device with classification capabilities. The radiologist interactively defined an ellipse encompassing the lesion, on the digital image, and activated the classification algorithm.

For mass lesions the CAD device automatically extracted quantitative features that characterized the mass encompassed by the ellipse. These features characterized the masses by their shape, definition of margins and speculation. Speculation was considered to be a structure composed of lines radiating from a centroid, rather than a sawtooth border of a lesion with a distinct margin. Therefore, this analysis could also be

applied to areas of architectural distortion, to focal asymmetries, to masses that appeared smoothly marginated, and to masses in which the margins were partially obscured.

For clusters of micro-calcifications, the CAD device automatically highlighted, in the first stage, an initial selection of potential micro-calcifications within the ellipse encompassing the cluster. The algorithm then allowed the radiologist to alter target selection, by modifying two detection filters in order to include only appropriate bright spots, which represent calcifications. The new selection of targets was updated in real time on the computer screen, and once the radiologist was satisfied with the selection of targets, the algorithm proceeded with automated extraction of features that characterize the cluster encompassed by the ellipse. For clusters two groups of features are automatically extracted by the computer as described in detail elsewhere [15]. The features in the first group reflect the shape, size and brightness of the individual micro-calcifications and those in the second group reflect the distribution of the calcifications within the cluster and cluster geometry.

Based on the extracted quantitative features, the classification algorithm, in the second stage, automatically generates descriptors taken from the BI-RADS lexicon, reflecting the appearance of the lesions in the mammogram. Descriptors that illustrate the appearance of a benign lesion are generated when low numerical values are obtained for the extracted features and descriptors that illustrate malignant lesions are generated when high values are obtained for the extracted features. The cut-point values for defining high and low numerical values for each of the extracted features were determined by the use of a separate training database with proven pathology results, which did not include any of the 109 cases described in the present study. The training database consisted of 500 cases of mammographically detected lesions with proven pathology, that were retrospectively collected from the archives of four other university-affiliated facilities, not including the facility from which the study cases were obtained. This database consisted of 289 mass lesions (161 malignant, 128 benign) and 211 clusters of micro-calcifications (94 malignant, 117 benign).

Figure 1 displays the descriptors that are generated by the classification algorithm for mass lesions based on their shape, margins and spiculation. The descriptors highlighted in white are generated for features with low values and describe the appearance of benign masses, while those highlighted in grey are generated for features with high values and describe the appearance of malignant masses.

The CAD-generated BI-RAD descriptors are displayed to the user for further assessment of the finding, as displayed in the example of a malignant mass in Figure 2.

For clusters, two sets of descriptors are generated by the CAD device. Figure 3 describes the first set of descriptors that are generated by the classification algorithm based on the appearance of individual calcifications in the lesion. Figure 4 describes the second set of descriptors that are based on the distribution of the calcifications within the cluster. The descriptors highlighted in white are generated for features with low values and describe the appearance of benign clusters, while those highlighted in grey are generated for features with high values and describe malignant masses.

A lesion is often assigned a combination of descriptors, some reflecting a benign appearance and some reflecting a malignant appearance, and then the resulting course of action is still to be defined. A paradigm was developed to assist the radiologist in assigning a level of suspicion, based on the computer generated descriptors. According to the paradigm a mass was considered benign if there was no evidence of

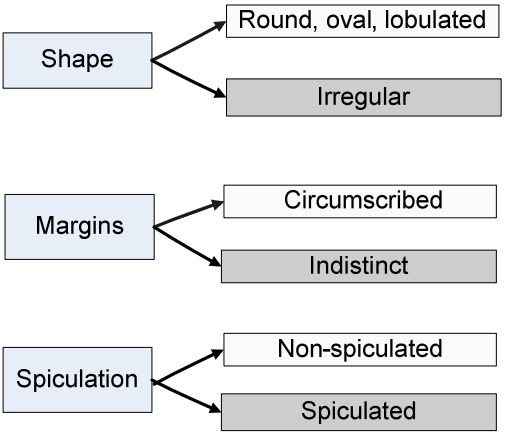

Fig. 1. BI-RADS descriptors generated by the CAD device for masses, based on their shape, margins and spiculation

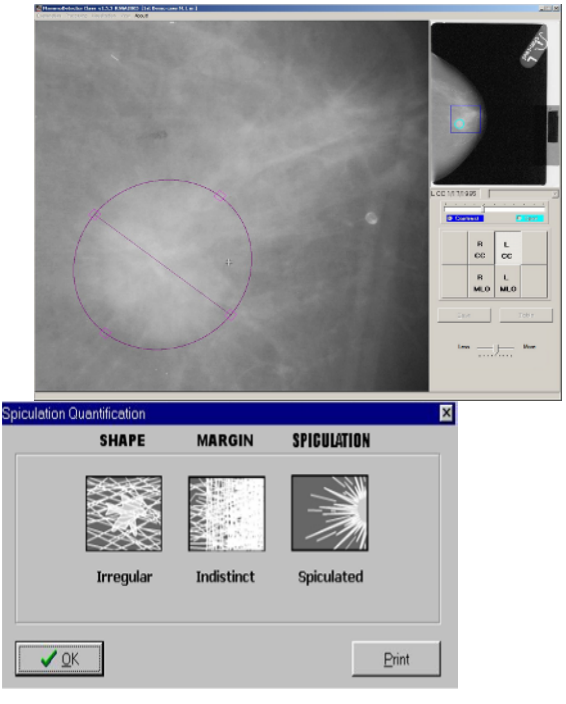

Fig. 2. The CAD-generated BI-RAD descriptors displayed to the user for a malignant mass

spiculation or if the mass was rounded and well circumscribed. Otherwise the mass was assigned a high level of suspicion and considered malignant. According to the paradigm developed for clusters, a cluster was considered benign, if the calcifications were not

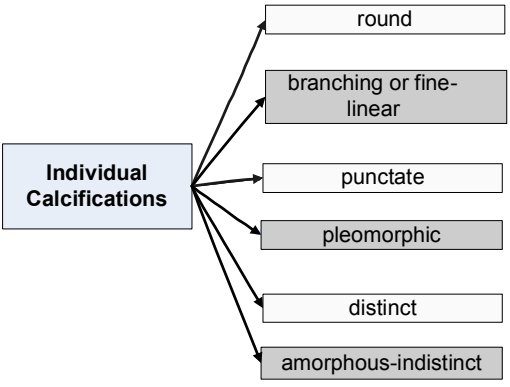

Fig. 3. BI-RADS descriptors generated by the CAD device for clusters, based on the appearance of individual calcifications

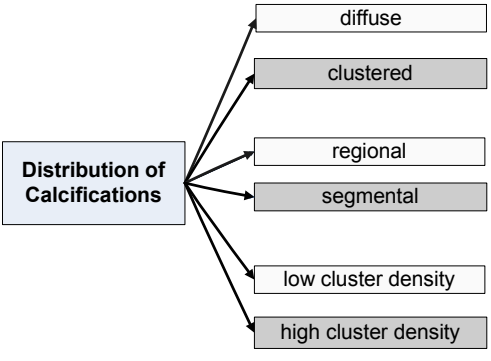

Fig. 4. BI-RADS descriptors generated by the CAD device for clusters, based on the distribution of the calcifications in the cluster

tightly clustered or if all the other 5 descriptors for the cluster reflected typically benign calcifications. In the latter category the calcifications are **not** amorphous, **not** branching **nor** fine-linear, **not** pleomorphic in shape, **not** segmental in distribution and the cluster is **not** of high density. Otherwise the cluster was considered malignant.

The advanced classification scheme generated descriptors characterizing all the 113 BI-RADS category 4A lesions included in the present study and the level of suspicion assigned by the paradigm based on the computer generated descriptors was compared with the pathology outcome of each lesion.

3 Results

According ACR BI-RADS suggestions, in all Category 4A lesions, biopsy should be considered, and the patient and her physician should make an informed decision on the ultimate course of action. All the BI-RADS Category 4A lesions, in this study were

forwarded for biopsy. Table 1 displays the results of the conventional interpretation versus the pathology outcome. As can be realized from Table 1, the conventional interpretation resulted for BI-RADS Category 4A lesions, in a Sensitivity of 100%, a Specificity of 0%, a Positive Predictive Value (PPV) of 8.8% and an overall accuracy of 8.8%.

Table 1. Results of the conventional interpretation of the BI-RADS Category 4A lesions

		Conventional Interpretation		
		+	-	Total
Pathology Results	+	10	0	10
	-	103	0	103
	Total	113	0	113

Table 2 displays the results of the computerized analysis, based on the CAD-generated descriptors, versus the pathology outcome. The paradigm, based on the CAD-generated BI-RADS descriptors, deemed malignant, all the 10 malignant BI-RADS Category 4A lesions, included in the study, yielding a sensitivity of 100% for that category. Of the 103 benign lesions, the computerized descriptors correctly classified 38 benign cases, yielding a specificity of 36.9% for BI-RADS Category 4A lesions. Of the 15 benign masses in the BI-RADS 4A Category, the paradigm deemed 12 masses benign, yielding a specificity of 80% for masses. Of the 88 benign clusters in the BI-RADS 4A Category, the paradigm deemed 26 clusters benign, yielding a specificity of 30% for clusters. The paradigm, based on the CAD-generated BI-RADS descriptors, yielded a Positive Predictive Value (PPV) of 13.3% and an overall accuracy of 42.5%, for the BI-RADS 4A Category lesions.

Figure 5 displays the performance of the paradigm based on the BI-RADS descriptors derived from the computer extracted quantitative features, compared to the results of the conventional assessment. This figure demonstrates the increase in the PPV and in the accuracy of diagnosis, caused by the use of the classification scheme, without any loss of sensitivity.

Table 2. Results of the computerized analysis, based on the CAD-generated descriptors, versus the pathology outcome. In this analysis the lesions were considered benign or malignant according to the outcome of the paradigm, using the BI-RADS descriptors.

		Computerized analysis based on CAD-generated descriptors		
		+	-	Total
Pathology Results	+	10	0	10
	-	65	38	103
	Total	75	38	113

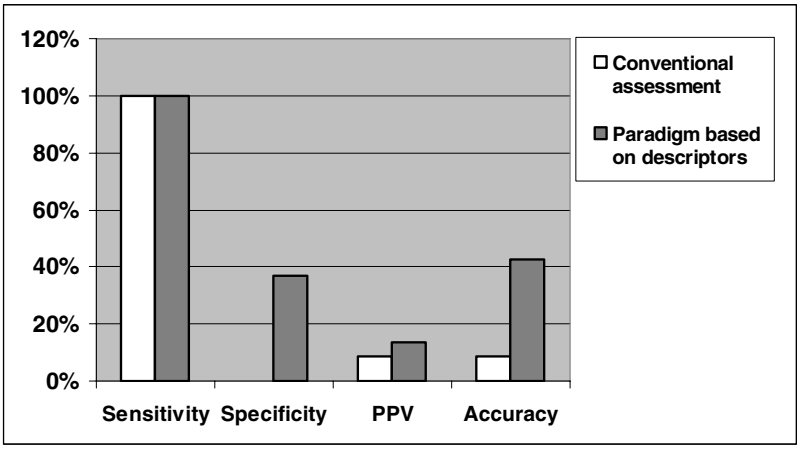

Fig. 5. The performance of the computerized analysis, compared to the conventional assessment

4 Conclusion

BI-RADS category 4A includes findings with the lowest level of suspicion, for which interventional procedures are, nevertheless still recommended. A very high percentage of biopsies performed in this category, results in a benign outcome. This study was performed to explore the hypothesis that computerized classification of the lesions in this category can reduce the number of unnecessary biopsies without affecting the sensitivity of diagnosis.

The use of a computerized analysis, based on BI-RADS descriptors generated by the CAD device, for BI-RADS 4A lesions, significantly increased the accuracy of diagnosis, from 8.8% to 42.5%, compared to conventional interpretation. The paradigm developed to assist the radiologist in establishing a course of action, based on the computer generated descriptors, eliminated 36.9% of unnecessary biopsies without decreasing the sensitivity. The paradigm for interpreting a finding based on these descriptors may well assist the radiologist in the complex task of assessing BI-RADS category 4A lesions.

References

1. Warren-Burhenne, L.L., Wood, S.A., D'Orsi, C.A., Feig, S.A., Kopans, D.B., O'Shaughnessy, K.F., Sickles, E.A., Tabar, L., Vyborny C.J. and Castellino, R.A.: Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology (2000) 215:554-562.
2. Ikeda, D.M., Birdwell, R.L., O'Shaughnessy, K.F., Sickles E.A. and Brenner R.J.: Computer-aided detection output on 172 subtle findings on normal mammograms previously obtained in women with breast cancer detected at follow-up screening mammography. Radiology, (2004) 230:811-819.

3. Roque, A.C. and Andre, T.C.: Mammography and computerized decision systems: a review. Ann. NY Acad. Sci. (2002) 980:83-94
4. Gur, D., Sumkin, J.H., Rockette, H.E., Ganott, M., Hakim, C., Hardesty, L., Poller, W.R., Shah, R. and Wallace. L.: Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. J. Natl. Cancer Inst. (2004) 96:185-190.
5. Doi, K., MacMahon, H., Katswragawa, S.,. Nishikawa, R.M, and Jiang, W.: Computer-aided diagnosis in radiology: potential and pitfalls. Eur. J. Radiol. (1999) 31:97-109
6. Lederman, R., Leichter, I., Novak, B., Bamberger, P., Fields, S., Buchbinder, S.: Stratification of mammographic cases by BIRADS category with the assistance of a CAD system. European Radiology. 13: (2003) 347-353.
7. Sickles, E.A.: Mammographic features of "early" breast cancer. Am J. Roengenol. (1984) 143:461-464 .
8. Kopans, D.B.: The positive predictive value of mammography. Am J. Roengenol. (1992) 158:521-526.
9. Fu, J.C., Lee, S.K., Wong, S.T. , Yeh, J.Y., Wang, A.H., Wu, H.K.: Image segmentation feature selection and pattern classification for mammographic microcalcifications. Comput Med Imaging Graph. (2005) 9:419-29.
10. Papadopoulos A., Fotiadis, D.I. , Likas, A.: Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines. Artif Intell Med. (2005) 34:141-50.
11. Wei, L., Yang, Y., Nishikawa, R.M., Jiang, Y.: A Study on Several Machine-Learning Methods for Classification of Malignant and Benign microcalcifications. IEEE Trans Med Imaging. (2005) 24:371-380
12. Burnside E.S., Sisney, G.A., Rubin, D.L., Ochsner, J.E., Fowler, K.: The Ability of Microcalcification Descriptors in the BI-RADS 4th Edition to Stratify the Risk of Malignancy. In: Proc. of 91th RSNA, November 27, 2005, Chicago, U.S.A.
13. Fields, S., Lederman, R., Buchbinder, S., Novak, B., Sklair, M., Bamberger, P., Leichter, I.: Improved mammographic accuracy with CAD assisted ranking of lesions. In: Proc. of the 6th IWDM, June 2002, Bremen, Germany
14. Fields, S., Leichter, I., Lederman, R., Buchbinder, S., Novak, B., Sklair-Levy, M., Sperber F., Bamberger, P.: Improving the Performance of Mammographic Assessment by the use of an Advanced Two- tiered CAD/CAC System.. In: Proc. of the 7th IWDM, June 2004, Charlotte, NC, U.S.A.
15. Leichter, I., Lederman, R., Bamberger, P., Novak, B., Fields, S., Buchbinder, S.: The use of an interactive software program for quantitative characterization of microcalcifications on digitized film-screen mammograms. Invest. Radiol. (1999) 34:394-400.