Development and Validation of a Fully Automated System for Detection and Diagnosis of Mammographic Lesions

Paola Casti[1*], Arianna Mencattini[1], Marcello Salmeri[1],
Antonietta Ancona[2], Fabio Mangieri[2], and Rangaraj M. Rangayyan[3]

***Abstract*— We present a comprehensive and fully automated system for computer-aided detection and diagnosis of masses in mammograms. Novel methods for detection include: selection of suspicious focal areas based on analysis of the gradient vector field, rejection of oriented components of breast tissue using multidirectional Gabor filtering, and use of differential features for rejection of false positives (FPs) via clustering of the surrounding fibroglandular tissue. The diagnosis step is based on extraction of contour-independent features for characterization of lesions as benign or malignant from automatically detected circular and annular regions. A new unified 3D free-response receiver operating characteristic framework is introduced for global analysis of two binary categorization problems in cascade. In total, 3,080 suspicious focal areas were extracted from a set of 156 full-field digital mammograms, including 26 malignant tumors, 120 benign lesions, and 18 normal mammograms. The proposed system detected and diagnosed malignant tumors with a sensitivity of 0.96, 0.92, and 0.88 at, respectively, 1.83, 0.46, and 0.45 FPs/image, with two stages of stepwise logistic regression for selection of features, a cascade of Fisher linear discriminant analysis and an artificial neural network with radial basis functions, and leave-one-patient-out cross-validation.**

I. INTRODUCTION

Computer-aided detection and diagnosis (CAD) of breast cancer offers a means to improve the efficiency of mammography and reduce breast cancer mortality [1]. Demands for reducing the number of unnecessary breast biopsies and obtaining higher sensitivity of detection of malignant tumors in mammographic screening indicate the urgent need of effective and comprehensive computerized solutions for mammography. Despite the number of successful attempts reported in the literature [1], [2], many challenges are still to be tackled for increased performance levels and transparency of the automated decision process. Significant records of research can be found dealing with detection [3]–[5] or diagnosis [6]–[9] of masses as independent tasks, but not many studies have addressed the detection and diagnosis stages in cascade [10], [11]. Characterization of masses as benign lesions or malignant tumors has been performed by using manually selected regions of interest and/or via accurate segmentation of the lesions. Hence, CAD remains a challenging task, especially due to the presence of lesions with ill-defined or obscured margins [8].

[1]*Dept. of Electronic Engineering, University of Rome Tor Vergata, Italy*
[2]*Diagnostic Radiology Unit, San Paolo Hospital of Bari, Italy*
[3]*Dept. of Electrical and Computer Engineering, Schulich School of Engineering, University of Calgary, Alberta, Canada*
[*]*Correspondence: casti@ing.uniroma2.it*

In this work, we address the issues introduced above by proposing a novel comprehensive approach for automatic detection and diagnosis of masses in a realistic scenario of a three-class space environment, i.e., in the presence of normal mammograms, benign lesions, and malignant tumors. This is achieved by solving two binary categorization problems in cascade: normal tissue versus mass candidates, followed by malignant tumors versus benign lesions. With this purpose, we introduce the following novel methodological aspects: 1) selection of mass candidates based on cardinality restrictions on the eigenvalues of the Hessian of the image; 2) rejection of physiological oriented structures of the breast parenchyma via analysis of the phase response of multidirectional Gabor filters; and 3) use of differential features for rejection of false-positive (FP) candidates that compare the detected suspicious focal areas with the surrounding fibroglandular tissue segmented by means of an iterative fuzzy c-means clustering algorithm [12]. The diagnosis stage in cascade represents an additional and important novelty of this work, being achieved by means of contour-independent features that do not require segmentation of the lesions [13], [14]. Such features are extracted from automatically detected circular and annular regions. Finally, we provide a unified 3D free-response receiver operating characteristic (FROC) framework for evaluation and analysis of the final results.

II. DATASET OF MAMMOGRAMS

A total of 156 full-field digital mammograms (FFDMs) were collected from 88 patients at the San Paolo Hospital of Bari, Italy, using the Senograph 2000D ADS 17.3 from GE Medical Systems. Informed consent for anonymous use of sensitive data for scientific purposes was obtained from all patients. The images have a spatial resolution of 94 μm and gray-scale resolution of 12 bits/pixel. The dataset is composed of 76 craniocaudal (CC) and 80 mediolateral-oblique (MLO) views, including 18 normal mammograms and 138 mammograms with masses. Based on biopsy results, out of the 146 lesions in the dataset, 26 were labeled as malignant tumors and 120 as benign lesions. The contours of the masses, used as ground truth for detection, were manually annotated and validated by two radiologists specialized in mammography (see the example in Fig. 1(a)).

III. METHODS

The proposed CAD scheme involves four main hierarchical modules: preprocessing of mammograms, detection of suspicious focal areas, rejection of falsely detected regions,

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

and diagnosis of malignant tumors. Two supervised classifiers of the mass candidates in cascade, one for detection and one for diagnosis, are applied after the extraction of specifically designed features.

A. Preprocessing of Mammograms

The original images were downsampled to the spatial resolution of 200 μm and enhanced using a look-up table (LUT). The breast region was extracted by applying previously developed methods [15], [16] for detection of the breast-skin line, the nipple, and the pectoral muscle (only for MLO views). A disk of radius 50 pixels (10 mm) centered at the position of the nipple was used to mask the nipple-areolar complex. The peripheral fat was masked by means of a ribbon of width 40 pixels (8 mm). An example of the obtained breast region is shown in Fig. 1.

B. Detection of Suspicious Focal Areas

Suspicious focal areas on the mammogram being processed were detected initially by analysis of the gradient vector field (GVF) [12], [16] and cardinality restrictions on the eigenvalues of the Hessian, as follows.

1) A Gaussian filter ($\sigma = 12$ pixels) and a box filter of size 20 pixels were applied to retain low-frequency information related to the cores of the masses.
2) The Hessian, H, embedding the second partial derivatives of the smoothed image I_s, was computed as

$$H = \begin{vmatrix} \frac{\partial^2 I_s}{\partial x^2} & \frac{\partial^2 I_s}{\partial x \partial y} \\ \frac{\partial^2 I_s}{\partial y \partial x} & \frac{\partial^2 I_s}{\partial y^2} \end{vmatrix}, \tag{1}$$

and the corresponding eigenvalues, $\lambda_1(x,y)$ and $\lambda_2(x,y)$, were evaluated at each pixel.
3) A set of suspicious focal areas having $\Lambda(x,y) = \lambda_1(x,y)\,\lambda_2(x,y) > 0$ and $\lambda_1(x,y) < 0$ was selected, in order to retain only nodes of convergent GVF, as shown in Fig.1(b) and (c).
4) The selected focal areas were rank-ordered based on the average value of $\Lambda(x,y)$, μ_Λ, for each region, up to a maximum of 20 candidates, and the center, $\mathbf{c_f}$, of the convex polygon enclosing each area was computed. The centers and rank values are shown in Fig.1(c).

C. Rejection of Falsely Detected Regions

Gabor filters were used to reject falsely detected regions based on their directionality. Given the real Gabor kernel, $g(x,y)$, oriented at $-\pi/2$, as

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \cos\left(2\pi\frac{x}{\tau}\right), \tag{2}$$

where the parameters σ_x, σ_y, and τ were derived using the design rules described by Ayres and Rangayyan [17], a set of 18 filters equally spaced over the angular range $(-\pi/2, \pi/2]$ was applied to derive a magnitude response, $M(x,y)$, shown in Fig. 1(d), and orientation field, $\Phi(x,y)$, by selecting the filter with the highest response at each pixel. The number of orientations in each detected area was computed and

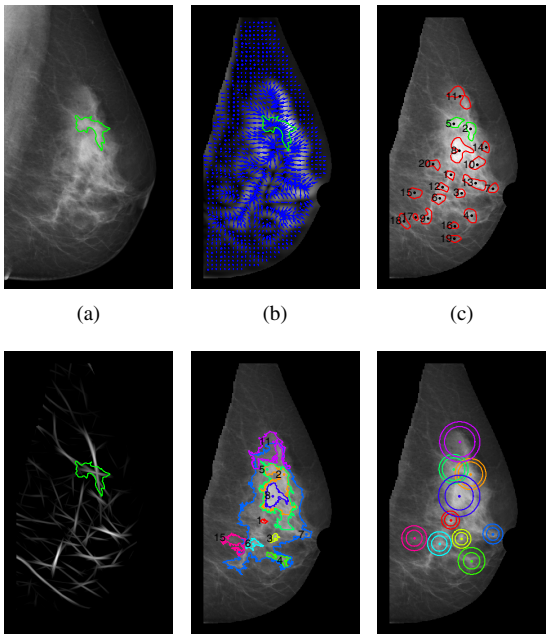


Fig. 1. (a) MLO mammogram including a lesion. The green contour indicates the region annotated by the radiologist. (b) GVF. (c) The 20 suspicious focal areas automatically detected using the eigenvalues of the GVF. The marked points indicate the centers of the lesions used for evaluation of results and ranked in descending order of μ_Λ. (d) Gabor magnitude response. (e) Examples of the surrounding fibroglandular tissue areas segmented via iterative fuzzy c-means clustering. (d) Automatically extracted circular and annular regions used for the diagnosis task.

mass candidates having less than six different orientations were rejected. An iterative fuzzy c-means algorithm [12] was applied to segment the fibroglandular tissue surrounding each remaining candidate. Examples of surrounding regions are illustrated in Fig. 1(e). Such regions, together with the suspicious focal areas previously detected, were used to extract a set of features for reduction of false mass candidates. The differences between the measures computed for each detected focal area, indicated by the subscript f, and the related measures for the surrounding tissue, indicated by the subscript t, were computed by obtaining a set of seven differential features, i.e., area (A and ΔA), compactness (ΔC_p), eccentricity (ε and $\Delta\varepsilon$), distance between centers (d), dispersion of orientation ($\Delta\sigma_\Phi$), weighted dispersion of orientation ($\Delta\sigma_w$), and dispersion of radii (σ_r), as follows:

$$\Delta A = |A_t - A_f| / (A_t + A_f), \tag{3}$$
$$\Delta C_p = \left|4\pi A_t/P_t^2 - 4\pi A_f/P_f^2\right|, \tag{4}$$
$$\Delta\varepsilon = |\varepsilon_t - \varepsilon_f|, \tag{5}$$
$$d = ||\mathbf{c_t} - \mathbf{c_f}||, \tag{6}$$
$$\Delta\sigma_\Phi = |std\left[\Phi_t(x,y)\right] - std\left[\Phi_f(x,y)\right]|, \tag{7}$$
$$\Delta\sigma_w = |std\left[(\Phi_t \circ M_t)_{x,y}\right] - std\left[(\Phi_f \circ M_f)_{x,y}\right]|, \tag{8}$$
$$\sigma_r = std\left[||B_t(x,y) - \mathbf{c_f}||\right] / mean\left[||B_t(x,y) - \mathbf{c_f}||\right], \tag{9}$$

where P and and $B(x,y)$ indicate, respectively, the perimeter and the boundary pixels of the fibroglandular tissue region. The symbol $\circ$ denotes the entrywise product. In addition, for

each candidate, μ_Λ and the normalized values for the maximum and standard deviation of the values of μ_Λ of all the detected candidates on each mammogram were computed.

D. Diagnosis of Malignant Tumors

Diagnostic classification of the detected regions as malignant tumors or benign lesions was performed using contour-independent features [13], [14] derived from automatically extracted circular and annular regions, X, in both the original mammogram and the Gabor magnitude response (see the regions in Fig. 1(f)). FPs from the detection stage also get passed on to the classification stage for analysis. The radius of each circular region was determined as the maximum radial distance of the center from the boundary of each focal area, while the width of the annular region was computed as described by Mudigonda et al. [11]. Quantification of radial and angular correlation was performed as follows

$$C = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} s_{ij} f_{ij}, \tag{10}$$

where s_{ij} is the radial or angular separation matrix and f_{ij} is a matrix of difference of pixel values. N is the number of concentric rings and angular sectors used. The angular and radial trends were computed as

$$T = \frac{1}{\#X}\sum_{i=1}^{N} \#\overline{X_i} < m_i, \tag{11}$$

were $\overline{X_i}$ is the complement of the angular or radial sectors analyzed and m_i is the related average intensity.

E. Pattern Classification and Cross-validation

Classification of the detected candidates as normal tissue or masses was achieved by means of Fisher-linear discriminant analysis (FLDA). The regions identified by FLDA as being masses, including FPs, were classified as malignant tumors or benign lesions using an artificial neural network with radial basis functions (ANN-RBF). The two sets of features used at each of the two classification steps were automatically selected via stepwise logistic regression using the training sets in each experiment. Leave-one-patient-out (LOO) and k-fold cross-validation were performed [18].

F. Performance Evaluation

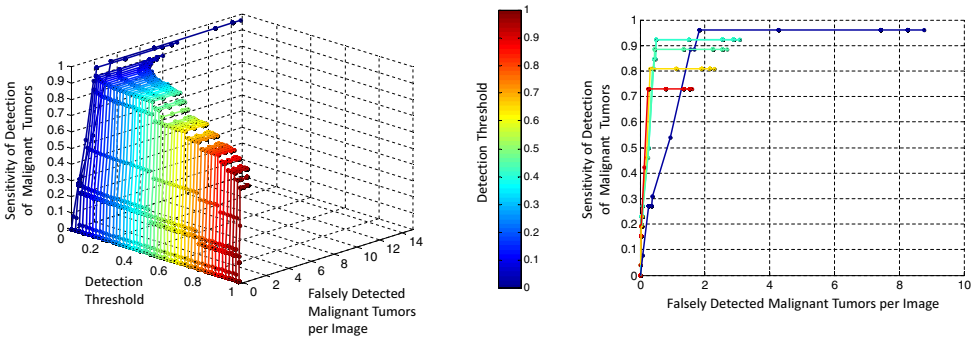
The performance of individual features was assessed in the context of detection (normal tissue versus mass candidates) and diagnosis (malignant versus benign candidates) by means of the area under the receiver operating characteristic (ROC) curve (A_z) [19]. Evaluation of combinations of the automatically selected features using different classifiers at each of the two stages in cascade was performed using ROC and FROC analysis. A mass candidate was considered as a true positive (TP) in detection and a TP in diagnosis if the center of the convex polygon enclosing the detected area was, respectively, within an annotated mass and within a malignant tumor. Additional TPs within the same annotated region, such as no. 2 and no. 5 in Fig. 1(c), were counted as one.

IV. RESULTS AND COMPARATIVE ANALYSIS

The individual capability of the proposed features for discrimination of masses from normal candidates was confirmed by A_z values (and standard error) up to 0.83 (0.02). In combination, the features selected by stepwise logistic regression achieved $A_z = 0.87$ (0.02) with FLDA and the leave-one-patient-out method. The corresponding sensitivities and FP values are summarized in Table I. The contour-independent features used for the diagnosis task were trained only on those automatically detected candidates that included benign lesions or malignant tumors. The highest value of A_z obtained individually was 0.75 (0.04). The features were combined using automatic feature selection in the LOO experiment and the ANN-RBF classifier. ROC analysis of the detected regions provided $A_z = 0.94$ (0.01) in the discrimination of malignant tumors from the rest of the candidates, including the normal tissue regions incorrectly classified as being masses in the preceding detection stage. Feature selection performed using the entire dataset provided similar results. Experiments with 2-fold and 10-fold cross-validation led to similar results as LOO at the detection stage; however, in malignant versus (benign+normal) classification, the A_z decreased, with the highest value of 0.79 ± 0.004 (average and standard deviation values for 100 trials of 10-fold cross-validation) obtained using two FLDA stages in cascade for detection and diagnosis. FROC analysis indicated 1.3 and 2.7 FPPI at 0.81 and 0.88 sensitivity, respectively, for malignant tumors. The reason for this deterioration with random splitting could be that the classifier may not have adequate malignant samples for training. A unified 3D FROC framework for detection and diagnosis of masses is presented in Fig. 2. The obtained sensitivity values for malignant tumors are shown in Fig. 2(a) for various combinations of threshold values of the output of the two classifiers for detection and diagnosis. The indicated numbers of falsely detected malignant tumors account for both normal tissue regions and benign lesions when incorrectly classified as being malignant. Some of the high-performance FROC curves for selected detection thresholds are illustrated in Fig. 2(b), from which the final results can be observed in detail. Table I summarizes the sensitivity and FP rates obtained. Overall, the system detected 96%, 92%, and 88% of malignant tumors with, respectively, 1.83, 0.46, and 0.45, FPs/image. The pro-

TABLE I
RESULTS OF FROC ANALYSIS FOR DETECTION OF MASSES AND DIAGNOSIS OF MALIGNANT TUMORS.

DETECTION	TPR (TP/# Masses)	0.80 (117/146)	0.86 (125/146)	0.90 (132/146)	0.95 (139/146)
	FP/Image (FP/# Images)	1.68 (262/156)	2.17 (338/156)	3.06 (478/156)	11.54 (1800/156)
DIAGNOSIS	TPR (TP/# Tumors)	0.81 (21/26)	0.88 (23/26)	0.92 (24/26)	0.96 (25/26)
	FP/Image (FP/# Images)	0.32 (50/156)	0.45 (70/156)	0.46 (72/156)	1.83 (285/156)



(a) (b)

Fig. 2. FROC analysis framework for global optimization of detection and diagnosis of malignant tumors for the dataset of 156 FFDM images. The detection threshold is indicated by the color map, while the diagnosis threshold varies along each curve. The number of falsely detected malignant tumors includes both normal tissue regions and benign lesions incorrectly classified as being malignant. (a) Sensitivity of detection of malignant tumors and falsely detected malignant tumors obtained by varying both detection and diagnosis thresholds. (b) Examples of FROC curves for selected detection thresholds.

cessing time of a mammogram is about 6 min using Matlab® R2014a. Taking into account the variety of databases and techniques used for evaluation of the results by the other reported methods, the results obtained in this study mostly outperform the state-of-the-art methods in the field [1]–[11] and our previous studies [13], [14]. As an example, in the recent work by Eltonsy et al. [5], sensitivities of 0.92, 0.88, and 0.81 at 5.4, 2.4, and 0.6 FPs/image, respectively, were reported using 270 CC mammograms with malignant tumors.

V. DISCUSSION AND CONCLUSIONS

A novel comprehensive scheme to detect malignant tumors in mammograms was presented in this study. Unlike previous work, we have addressed the integration of detection and diagnosis in a unified CAD scheme. A 3D FROC framework facilitates a realistic interpretation of results and allows the optimal setting of the operating point by the user. Most importantly, our system can detect and then diagnose malignant tumors without the requirement of accurate segmentation of their contours. This is important when dealing with lesions with obscured or ill-defined borders and with dense breasts. The obtained results indicate that detection and diagnosis can be combined together to increase the sensitivity of detection of malignant tumors while maintaining low FP rates. Future work will incorporate additional evaluation and testing using publicly available databases (e.g., DDSM [20]).

REFERENCES

[1] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *J. Comp. Med. Imaging Graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.

[2] J. Tang, R.M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 236–251, 2009.

[3] N. Karssemeijer and G.M. te Brake, "Detection of stellate distortions in mammograms," *IEEE Trans. Med. Imag.*, vol. 15, no. 5, pp. 611–619, 1996.

[4] N. Petrick, B. Sahiner, H.P. Chan, M.A. Helvie, S. Paquerault, and L.M. Hadjiisky, "Breast cancer detection: evaluation of a mass-detection algorithm for computer-aided diagnosis - experience in 263 patients," *Radiology*, vol. 224, no. 1, pp. 217–224, 2002.

[5] N.H. Eltonsy, G.D. Tourassi, and A.S. Elmaghraby, "A concentric morphology model for the detection of masses in mammography," *IEEE Trans. Med. Imag.*, vol. 26, no. 6, pp. 880–889, 2007.

[6] B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvie, and L.M. Hadjiiski, "Improvement of mammographic mass characterization using spiculation measures and morphological features," *Med. Phys.*, vol. 28, no. 7, pp. 1455–1465, Jul. 2001.

[7] A. Rojas-Domínguez and A.K. Nandi, "Development of tolerant features for characterization of masses in mammograms," *Comput. Biol. Med.*, vol. 39, no. 8, pp. 678–688, 2009.

[8] M. Elter and A. Horsch, "CADx of mammographic masses and clustered microcalcifications: a review," *Med. Phys.*, vol. 36, no. 6, pp. 2052–2068, 2009.

[9] A. Rojas-Domínguez and A.K. Nandi, "Toward breast cancer diagnosis based on automated segmentation of masses in mammograms," *Pattern Recogn.*, vol. 42, no. 6, pp. 1138–1148, 2009.

[10] W.E. Polakowski, D.A. Cournoyer, S.K. Rogers, M.P. DeSimio, D.W. Ruck, J.W. Hoffmeister, and R.A. Raines, "Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivative-based feature saliency," *IEEE Trans. Med. Imag.*, vol. 16, no. 6, pp. 811–819, 1997.

[11] N.R. Mudigonda, R.M. Rangayyan, and J.E.L. Desautels, "Detection of breast masses in mammograms by density slicing and texture flow-field analysis," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1215–1227, 2001.

[12] A. Mencattini and M. Salmeri, "Breast masses detection using phase portrait analysis and fuzzy inference systems," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 7, no. 4, pp. 573–583, 2011.

[13] P. Casti, A. Mencattini, M. Salmeri, A. Ancona, F. Mangieri, and R.M. Rangayyan, "Measures of radial correlation and trend for classification of breast masses in mammograms," in *Conf. Proc. IEEE Eng. Med. Biol. Soc. (EMBS)*, Osaka, Japan, 3-7 Jul. 2013.

[14] P. Casti, A. Mencattini, M. Salmeri, A. Ancona, F. Mangieri, and R.M. Rangayyan, "Design and analysis of contour-independent features for classification of mammographic lesions," in *Conf. Proc. IEEE e-Health and Bioeng. (EHB)*, Iasi, Romania, 21-23 Nov. 2013.

[15] P. Casti, A. Mencattini, M. Salmeri, A. Ancona, F.F. Mangieri, M.L. Pepe, and R.M. Rangayyan, "Estimation of the breast skin-line in mammograms using multidirectional Gabor filters," *Comput. Biol. Med.*, vol. 43, no. 11, pp. 1870–1881, 2013.

[16] P. Casti, A. Mencattini, M. Salmeri, A. Ancona, F.F. Mangieri, M.L. Pepe, and R.M. Rangayyan, "Automatic detection of the nipple in screen-film and full-field digital mammograms using a novel Hessian-based method," *J. Digit. Imaging*, vol. 26, no. 5, pp. 948–957, 2013.

[17] F.J. Ayres and R.M. Rangayyan, "Design and performance analysis of oriented feature detectors," *J. Electron. Imaging*, vol. 16, no. 2, pp. 12 pages, April 2007, article number 023007.

[18] R.O. Duda, P.E. Hart, and DG Stork, *Pattern Classification*, Wiley-Interscience, 2nd edition, 2001.

[19] "ROCKIT software," http://www.radiology.uchicago.edu.

[20] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W.P. Kegelmeyer, "The Digital Database for Screening Mammography," in *Proc. 5th Int. Workshop on Digital Mammography*, Medical Physics Publishing, Ed., 2001, pp. 212–218.