On Detecting Abnormalities in Digital Mammography

Waleed A. Yousef[1], Waleed A. Mustafa, Ali A. Ali, Naglaa A. Abdelrazek[2], and Ahmed M. Farrag[1]
[1]Faculty of Computers and Information, Helwan University, Egypt,
[2]Faculty of Medicine, Cairo University, Egypt.
Wyousef@Helwan.edu.eg, {ali.ahmed.saleh, waleedAMustafa, a.m.farrag}@gmail.com, NaglaAbdelrazek@Yahoo.com

***Abstract*—Breast cancer is the most common cancer in many countries all over the world. Early detection of cancer, in either diagnosis or screening programs, decreases the mortality rates. Computer Aided Detection (CAD) is software that aids radiologists in detecting abnormalities in medical images. In this article we present our approach in detecting abnormalities in mammograms using digital mammography. Each mammogram in our dataset is manually processed—using software specially developed for that purpose—by a radiologist to mark and label different types of abnormalities. Once marked, processing henceforth is applied using computer algorithms. The majority of existing detection techniques relies on image processing (IP) to extract Regions of Interests (ROI) then extract features from those ROIs to be the input of a statistical learning machine (classifier). Detection, in this approach, is basically done at the IP phase; while the ultimate role of classifiers is to reduce the number of False Positives (FP) detected in the IP phase. In contrast, processing algorithms and classifiers, in pixel-based approach, work directly at the pixel level. We demonstrate the performance of some methods belonging to this approach and suggest an assessment metric in terms of the Mann Whitney statistic.**

***Index Terms*—Classification, Detection, Image Processing, Digital Mammography, Breast Cancer, Computer Aided Detection (CAD).**

I. Introduction

Detection of breast cancer while it is still small and confined to the breast provides the best chance of effective treatment for women with the disease [1, 2]. Benefits of early detection include increased survival rate, increased treatment options and improved quality of life. Currently, there is insufficient knowledge about the causes of breast cancer for primary prevention strategies to reduce incidence in the population.

Causes of missed breast cancer on mammography can be secondary to many factors including those related to the patient (whether inherent or acquired), the nature of the malignant mass itself, poor mammographic techniques, or provider factors or interpretive skills of radiologists and oncologists (including perception and interpretation errors) [3].

Perception error occurs when the lesion is included in the field of view and is evident but is not recognized by the radiologist. The lesion may or may not have subtle features of malignancy that cause it to be less visible. Small nonspiculated masses, areas of architectural distortion, asymmetry, small clusters of amorphous or faint microcalcifications. All may be difficult to perceive [3].

Several factors may lead to misinterpretation, such as lack of experience, fatigue, or inattention. Misinterpretation may also occur if the radiologist fails to obtain all the views needed to assess the characteristics of a lesion or if the lesion is slow growing and prior images are not used for comparison [4, 3].

The implementation of Computer aided detection (CAD) systems will help to reduce the human errors that lead to missing breast carcinoma, either related to poor perception or interpretation errors. CAD could increase the sensitivity of mammography interpretation [5].

The American Food and Drug Administration (FDA) had approved the use of Computer Aided Devices (Detection or Diagnosis) in 1998; since then many CAD systems have been developed. Despite the availability of such systems all over the world, and in the U.S. in particular, they have no existence in many countries for their exaggerated price which ranges from 50,000$ to 175,000$ [6].

In retrospect, developing a CAD system that is affordable to all laboratories, and individual radiologists on their desktop, is of great value to the field of medical diagnosis.

This manuscript presents current attempts in our project towards detecting abnormalities in digital mammograms. Because of manuscript length limitations, we try to give a breadth-first exposition to our approach. Details on each particular issue should be assigned to subsequent publications. We focus our attention in the present article to detecting masses.

The working team comprises a multidisciplinary group of several backgrounds including statistics, computer science, and engineering, along with a trained, experienced and professional radiologist (10 years experience, 5000 mammogram / year). We collect the mammograms from two different institutions; all images are acquired from digital mammography.

The radiologist reads the screening digital mammograms and then mark the lesions in the images. The marked lesions are also tagged according to the different radiological lexicons and then categorized by the radiologist according to the "Breast Imaging Reporting and Data System" (BIRADS) scoring system. Table I is the description of the international BIRADS scoring system for diagnosis of breast lesions. We have implemented our

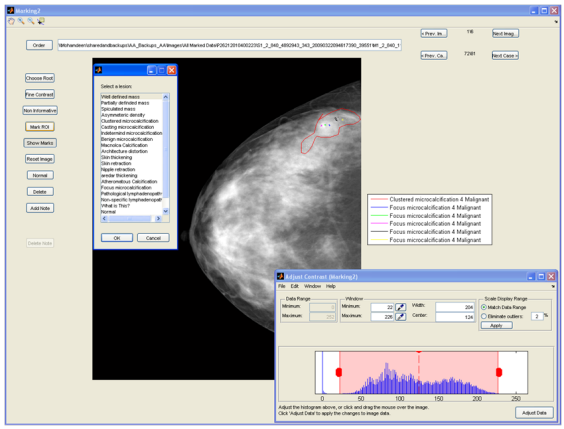


Fig. 1. A snapshot for a software used by the consultant radiologist to mark every lesion in a mammogram. The radiologist marks accurately the outline of a lesion with entering all related information, e.g., BIRAD and histological results

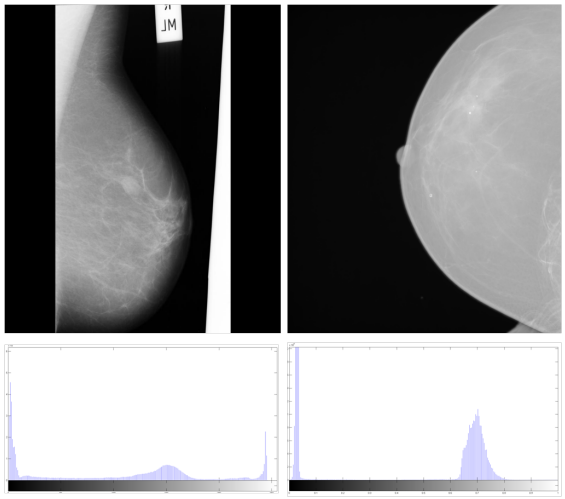


Fig. 2. A snapshot of a software used by the consultant radiologist to mark every lesion in a mammogram. The radiologist marks accurately the outline of a lesion and enters all related information, e.g., BIRADS and histological results

protocol for reading and marking by designing a software that facilitates the radiologist with labeling, marking, and attributing tools. Figure 1 is a snapshot of this software. All lesions were classified according to the BIRADS system, then a BIRADS score was assigned for each image [7]. All suspicious lesions classified as BIRADS 3, 4 and 5 are pathologically proven after core and vacuum needle biopsy.

Category 0	mammographic assessment is incomplete
Category 1	negative
Category 2	benign finding(s)
Category 3	probably benign finding(s)
Category 4	suspicious abnormality
Category 5	highly suggestive of malignancy

TABLE I
"BREAST IMAGING REPORTING AND DATA SYSTEM" (BIRADS) SCORING.

The rest of the paper is organized as follows. In section II we briefly explore the different two common approaches for designing a CAD, the pixel- vs. the region-based approaches. In this section we describe several methods available in the literature, that subscribe to the former approach. In section III, we discuss some technical issues related to the use of the FROC for assessing CADs. In Section IV we provide a comparative study among different methods in pixel-based design. In Section V we conclude the paper and discuss our current work in this project for producing an affordable CAD.

II. DESIGN

In this section, we introduce some of the methods used in our comparative study; the details can be found in the literature. As mentioned above, the methods leveraged here belong to the pixel-based category of CADs, in which each pixel is represented as a separate feature vector and is classified individually. This approach should be contrasted to the region-based approach which forms regions, e.g., by segmentation, then a feature vector is formed for each region. Afterwards, classification is carried out for a region as a whole. In the later approach, detection primarily takes place at the first step of segmentation; subsequent classification is for False Positive (FP) reduction. A good review that discusses both approaches is [8].

Data preprocessing is essential; it normalizes data for a particular method. One fact that is worth mentioning is that digital mammograms (as those used in our project) greatly abandon many preprocessing steps. Figure 2 displays analogue and digital mammograms, along with associated histogram. It is evident that those unpleasing artifacts, marks, and image noise in the analogue image have no existence in the digital mammogram; this resulted in a much "cleaner" histogram, in the sense that it has only two well-separated sub-histograms for background and breast. In effect, all preprocessing steps required for removing artifacts, noise, and marks are abandoned. Moreover, breast segmentation—which was a literature by itself for analogue mammography (see, e.g., [9] that covers some of this literature)—is no longer a great deal of research for digital mammography. The well-separated background and breast histograms of the digital mammography (Figure 2) renders simple segmentation methods as K-means clustering or Otsu's method (see, e.g., [10]) very successful.

A. IRIS Filter

As reported in [11], Kobatake's IRIS Filter [12, 13] is one of the highly ranked methods in detecting masses. We implement IRIS filter for comparison with other methods.

According to [13], the IRIS filter is a 2D filtering technique to enhance rounded, semi-rounded, and stellate opacities in an image, regardless of their contrast to the background. The main idea of the IRIS filter is to measure the convergence index on each pixel in the gradient orientation map of a mammogram to measure how strongly the gradient vectors point toward a

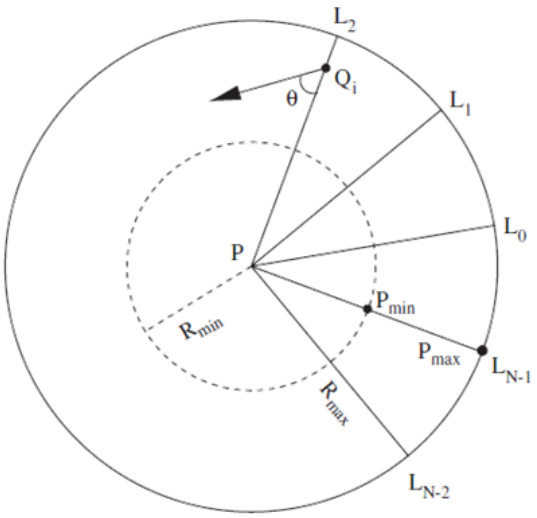


Fig. 3. Pixel of interest with associated radiating lines: how IRIS convergence index is calculated (Courtesy Kobatake)

pixel of interest. This is done by first calculating the gradient orientation map of the image using a 3×3 gradient operator. The gradient amplitude of the image is

$$|g(x,y)|^2 = |G_R(x,y)|^2 + |G_C(x,y)|^2,$$

where $G_R(x,y)$ is the row gradient, $G_C(x,y)$ is the column gradient. The orientation of the gradient is calculated as

$$\phi(x,y) = \tan^{-1} \frac{G_C(x,y)}{G_R(x,y)}.$$

The IRIS filter is then applied by examining each pixel of interest (x,y), taking N half lines around it in a circular manner (Figure 3). Then, for each half line, every pixel at a distance d, $R_{min} < d < R_{max}$, on the line is examined to calculate the convergence degree. The maximum C_i over a half line is defined as

$$C_i(x,y) = \max \frac{\int_P^{Q_i} \cos\theta(x,y) dQ}{\overline{PQ_i}}, \ 0 < \overline{PQ_i} < R_{max}$$

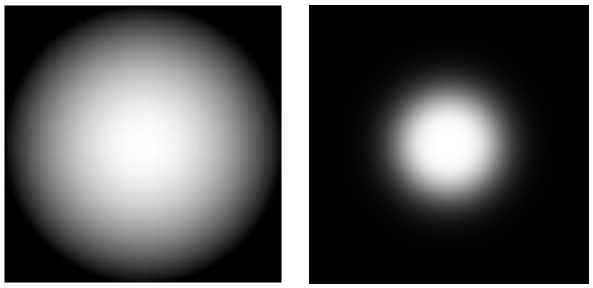
The output of the IRIS filter for the pixel P is the average over all half lines:

$$C(x,y) = \frac{1}{N} \sum_{i=0}^{N-1} C_i(x,y).$$

We tried $R_{min} = 5, 7, 10, 13, 15$ pixels; $R_{max} = 35$ pixels; $Z = 10, 8, 6, 4$ (a scaling factor to down-sample the images). The best configuration, giving an AUC (see Section III) of 0.883, is $R_{min} = 5, \ R_{max} = 3, \ Z = 10$.

B. Template Matching

Template Matching is a method that measures the correlation between a subregion of an image and a predefined template. Several templates exist in the literature such as the spherical template shown in Figure 4(a) (used in [14]), and the hyperbolic secant template shown in Figure 4(b) (used in [15]).



(a) Spherical Template (b) Hyperbolic Secant Template

Fig. 4. Templates

The spherical template is defined as

$$T(x,y) = \begin{cases} R^2 - x^2 - y^2, & \text{if } x^2 + y^2 < R^2 \\ 0, & \text{if } x^2 + y^2 \geq R^2 \end{cases},$$

where R is the template radius. The hyperbolic secant template is defined as

$$T(x,y) = sech(x+y).$$

We fixed the scaling factor Z at 10. We varied the radius of the template as: $R = 4\% \ldots 12\%$ (relative to the image hight). The multiscale analysis described in [14] has been implemented by changing the template size and then combining the scores (correlation) of different templates. The resultant score for each pixel is the maximum of all the scores given to the same pixel by different template sizes. The multiscale combination $(6\%, 8\%, 10\%)$ gave the best AUC (see Section III) of 0.886.

C. Featureless approach

The idea of this approach appears in some publications, e.g., for detecting masses [16] and microcalcifications [17]. The basic principle of this approach is to take the gray level of the pixel of interest along with its neighborhood and apply Support Vector Machine as a classifier. The approach is "featureless" from the sense that no features are extracted by processing. The gray levels of the pixel of interest along with its neighborhood are indeed the features. The same idea had existed in the StatLog project [18] for classifying areas of different crops from a multi-spectral satellite images.

We adopted the same idea; however, we manipulated many other different tuning and complexity parameters. The details of this approach is out of the scope of the present article, and is deferred to separate publication.

III. Assessment

The common practice in the literature is to assess CADs in terms of the Free-response Receiver Operating Characteristics (FROC) [19]. FROC is a plot of the total number of detected abnormalities (normalized by the number of images)—therefore called True Positive Fraction (TPF)—versus the average number of False Positive (FP) markers per image.

Assessment in terms of the FROC for the pixel-based methods (like those presented in the current paper), is not a

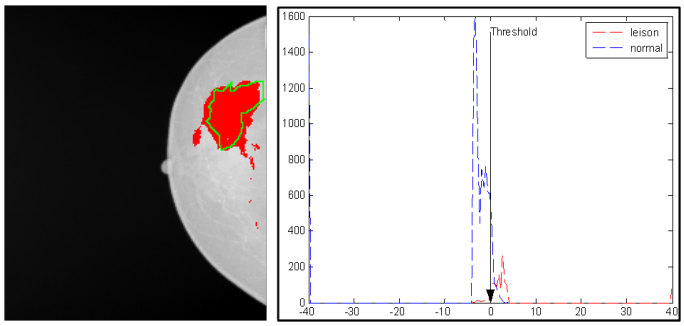


Fig. 5. Left: an image with radiologist annotation (green boundary) and detected region is labeled in red. Right: the histograms of the image of scores using one of the methods. The larger the separation between the two histograms the higher the AUC of this image. The AUC for this image is 0.97.

straightforward task. This is because there is no agreement in the literature on two main issues. First, what is the criterion on which we define a detected region to be TP. Many definitions exist in the literature, and the performance of a method can vary significantly by changing the criterion. Second, what is the moving threshold that increases the TP and FP? Is it the threshold working on the pixel scores or something else? We are preparing another manuscript to answer those questions. Meanwhile, we propose another criterion for assessment in terms of the Mann Whitney statistic, named differently the Area Under the Curve (AUC). The AUC of a score image (probability images) is defined as

$$\widehat{AUC} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I_{(x_i < y_j)},$$

$$I_{(c)} = \begin{cases} 1, & \text{if c is true} \\ 0, & \text{if c is false} \end{cases}$$

where x is the set of scores of the normal region, y is the set of scores of the malignant region(s), and n_1 and n_2 are the respective number of pixels. This is a measure of the separability of the two sets of scores. The closer the AUC to a value of 1 the more the separability between the two sets of scores.

Figure 5 demonstrates our definition above. The image on the left of the figure has a malignant region marked in green by the radiologist, along with the detected pixels colored in red. The right image shows the histograms of the two sets of scores given by a particular method to each region. The detected pixels on the left are obtained by setting the threshold value (indicated on the right) to zero. The AUC of this image is 0.97. This is a single number expressing how the pixel scoring of a particular method separates the malignant from the normal regions. More explanation is provided in Section IV when we discuss the results.

IV. Results and Discussion

We applied the three methods introduced in Section II to a dataset consisting of 50 malignant cases and 100 normals. Each method has its own tuning parameters, as explained in Section II. The figure of merit for assessment is the average AUC (defined in Section III), where the average is taken over images. The average AUC for each of the three methods (IRIS, template matching, and featureless) are 0.866, 0.883, and 0.953 respectively.

It is worth mentioning that for the featureless approach, when we tried several complexity parameters, we selected the "best" parameters in a leave-one-case-out resampling paradigm, in which we train the model on all, but one case, then test on this left-out case. Therefore, a case with m images contributes with m AUC values. The AUC itself for one image is obtained as a Mann Whitney statistic (as explained in Section III).

The images in Figures 6–7 provide a better understanding for the assessment part. Every two consecutive rows in these figures concern only one image from our dataset; therefore, these two figures illustrate four different images. For example, the first row is consisted of the original image with the radiologist annotation (first image), along with the pixel score images provided by each of the three methods. The three methods give high score to the malignant region; however, the featureless method strongly illuminate that region. The AUC values provided at the bottom of the second row objectively assesses how each method does separate the two regions. For this particular image, the IRIS filter, template matching, and the featureless approach give AUC of 0.81, 0.91, 0.99 respectively. It is quite informative to compare these numbers to the AUC of the original image, i.e., 0.96. The later represents how the two regions in the original image (the plain gray level image without any processing) are already separated. This conveys how well each method does really provide to enhance this particular image.

The second row represents the regions detected by each method, obtained by moving a threshold value until obtaining a "satisfactory" detected region. The pixels above this threshold value are colored in red; some false positives are obvious. It is worth mentioning that the threshold that works well for both IRIS and template matching is almost constant for each method.

V. Future and Current Work in Progress

As mentioned in the introduction, although there are some commercial CADs available in the market, they are very expensive for many laboratories. Because of their very high prices, CADs, in some countries, have no existence in any laboratory. Our objective in this project is to produce a Computer Aided Detection (CAD) system that is: 1) affordable (for laboratories and radiologists) and 2) outperforming.

For achieving the first goal, our CAD, which is still under design, is planned to be a software without special hardware requirements. Moreover, we are planning to produce the CAD in the form of a Dynamic Linked Library (DLL), which is a set of software routines, so that they can be called from any other simple image viewer.

For achieving the second goal, we are working in our lab. at re-producing many of the available methods in the literature that proved to be superior to others. This would allow us to compare our own designed method to those other methods,

and to combine the best methods in a more general multiple classifier system (MCS) framework for obtaining an overall outperforming CAD. By the time of writing this manuscript we managed to produce a first version of a combined classifier that has better performance than any other individual method tested in this project. The methodology of combining those classifiers (or CADs) to produce a new outperforming CAD is to be discussed in future publications.

VI. Acknowledgment

This research is funded by the Information Technology Industry Development Agency (ITIDA), Egypt, under grant number ARP2009.R6.3. The authors, are grateful to Nicholas Petrick, of the U.S. Food and Drug Administration (FDA), for fruitful discussions related to this manuscript.

References

[1] J. Ferlay, F. Bray, P. Pisani, and D. Parkin, "Globocan 2002: Cancer incidence, mortality and prevalence worldwide." 2004.

[2] D. M. Parkin, F. Bray, J. Ferlay, and P. Pisani, "Global cancer statistics, 2002," *Ca-a Cancer Journal for Clinicians*, vol. 55, no. 2, pp. 74–108, 2005.

[3] R. M. Kamal, N. M. Abdel Razek, M. A. Hassan, and M. A. Shaalan, "Missed breast carcinoma; why and how to avoid?" *J Egypt Natl Canc Inst*, vol. 19, no. 3, pp. 178–94, 2007.

[4] S. Radhika, "Imaging techniques alternative to mammography for early detection of breast cancer," *TRCT*, pp. 1–120, 2005.

[5] M. Muttarak, S. Pojchamarnwiputh, and B. Chaiwun, "Breast carcinomas: why are they missed?" *Singapore Med J*, vol. 47, no. 10, pp. 851–7, 2006.

[6] F. M. Hall, "Breast imaging and computer-aided detection," *N Engl J Med*, vol. 356, no. 14, pp. 1464–1466, 2007.

[7] B. O. Anderson, C. H. Yip, S. D. Ramsey, R. Bengoa, S. Braun, M. Fitch, M. Groot, H. Sancho-Garnier, V. D. Tsu, and P. Global Hlth Care Systems Public, "Breast cancer in limited-resource countries: Health care systems and public policy," *Breast Journal*, vol. 12, no. 1, pp. S54–S69, 2006.

[8] J. S. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 2, pp. 236–251, 2009.

[9] D. Raba, A. Oliver, J. Marti, M. Peracaula, and J. Espunya, "Breast segmentation with pectoral muscle suppression on digital mammograms," *Pattern Recognition and Image Analysis, Pt 2, Proceedings*, vol. 3523, pp. 471–478, 2005.

[10] R. C. Gonzalez and R. E. Woods, *Digital image processing*, 2nd ed. Upper Saddle River, N.J.: Prentice Hall, 2002, rafael C. Gonzalez, Richard E. Woods. ill. ; 25 cm.

[11] A. Oliver, J. Freixenet, J. Marti, E. Perez, J. Pont, E. R. E. Denton, and R. Zwiggelaar, "A review of automatic mass detection and segmentation in mammographic images," *Medical Image Analysis*, vol. 14, no. 2, pp. 87–110, 2010.

[12] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized detection of malignant tumors on digital mammograms," *IEEE Trans Med Imaging*, vol. 18, no. 5, pp. 369–78, 1999.

[13] H. Kobatake and S. Hashimoto, "Convergence index filter for vector fields," *Ieee Transactions on Image Processing*, vol. 8, no. 8, pp. 1029–1038, 1999.

[14] G. M. te Brake and N. Karssemeijer, "Single and multiscale detection of masses in digital mammograms," *IEEE Trans Med Imaging*, vol. 18, no. 7, pp. 628–39, 1999.

[15] S. Morrison and L. M. Linnett, "A model based approach to edge detection and parameterisation," *Seventh International Conference on Image Processing and Its Applications*, no. 465, pp. 202–205 869, 1999.

[16] R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, and M. Roffilli, "A novel featureless approach to mass detection in digital mammograms based on support vector machines," *Physics in Medicine and Biology*, vol. 49, no. 6, pp. 961–975, 2004.

[17] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Trans Med Imaging*, vol. 21, no. 12, pp. 1552–63, 2002.

[18] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, *Machine learning, neural and statistical classification*, ser. Ellis Horwood series in artificial intelligence. New York ; London: Ellis Horwood, 1994.

[19] D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modeling, analysis, and validation," *Medical Physics*, vol. 31, no. 8, pp. 2313–2330, 2004.

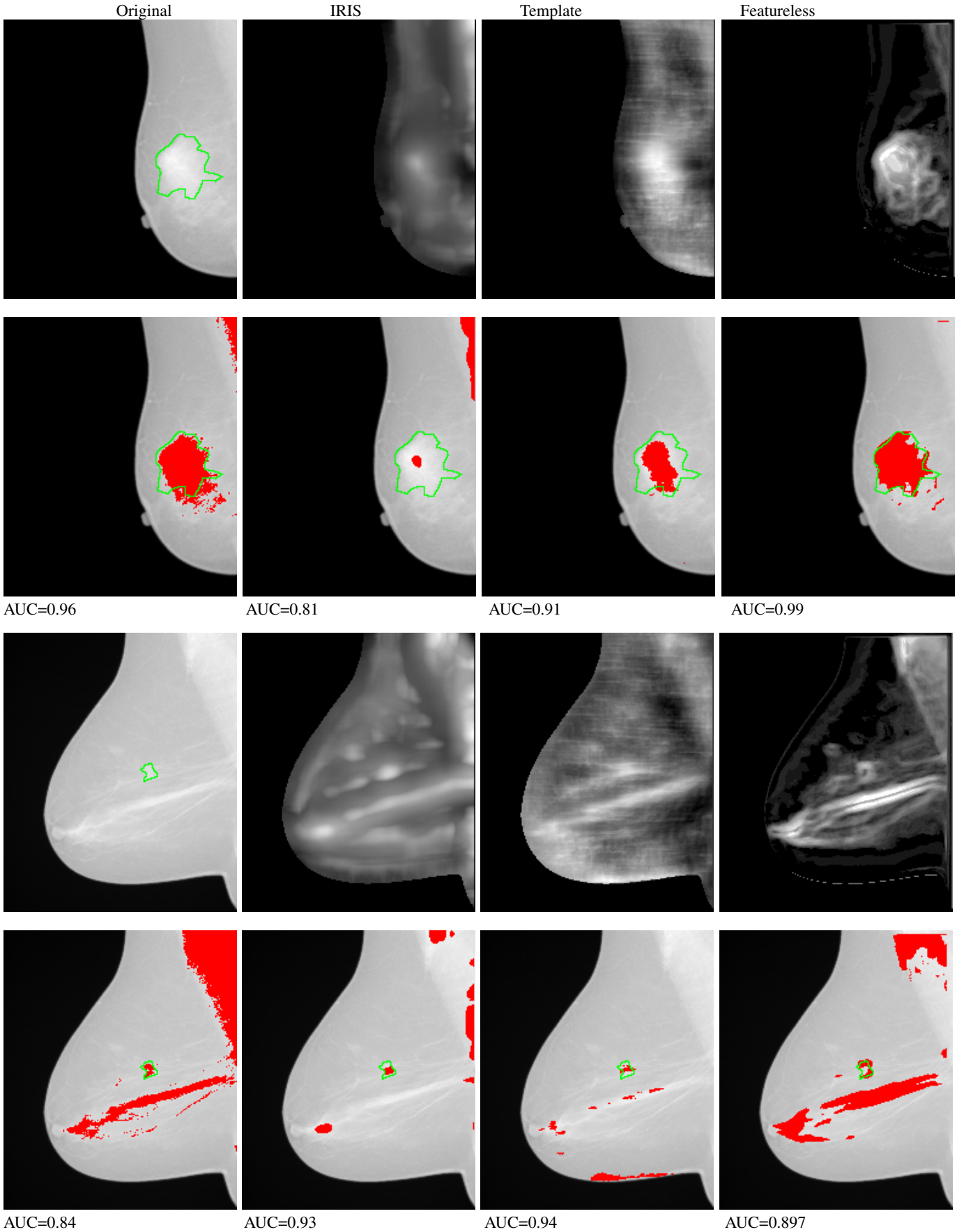


Fig. 6. Upper left is an image with radiologist's marking; the other three images in the same row are the scores given by different three methods. The second row is the detected regions by setting a threshold for the score image. Third and Forth rows are the results of another image. The AUC of each particular method is indicated at the bottom of each row.

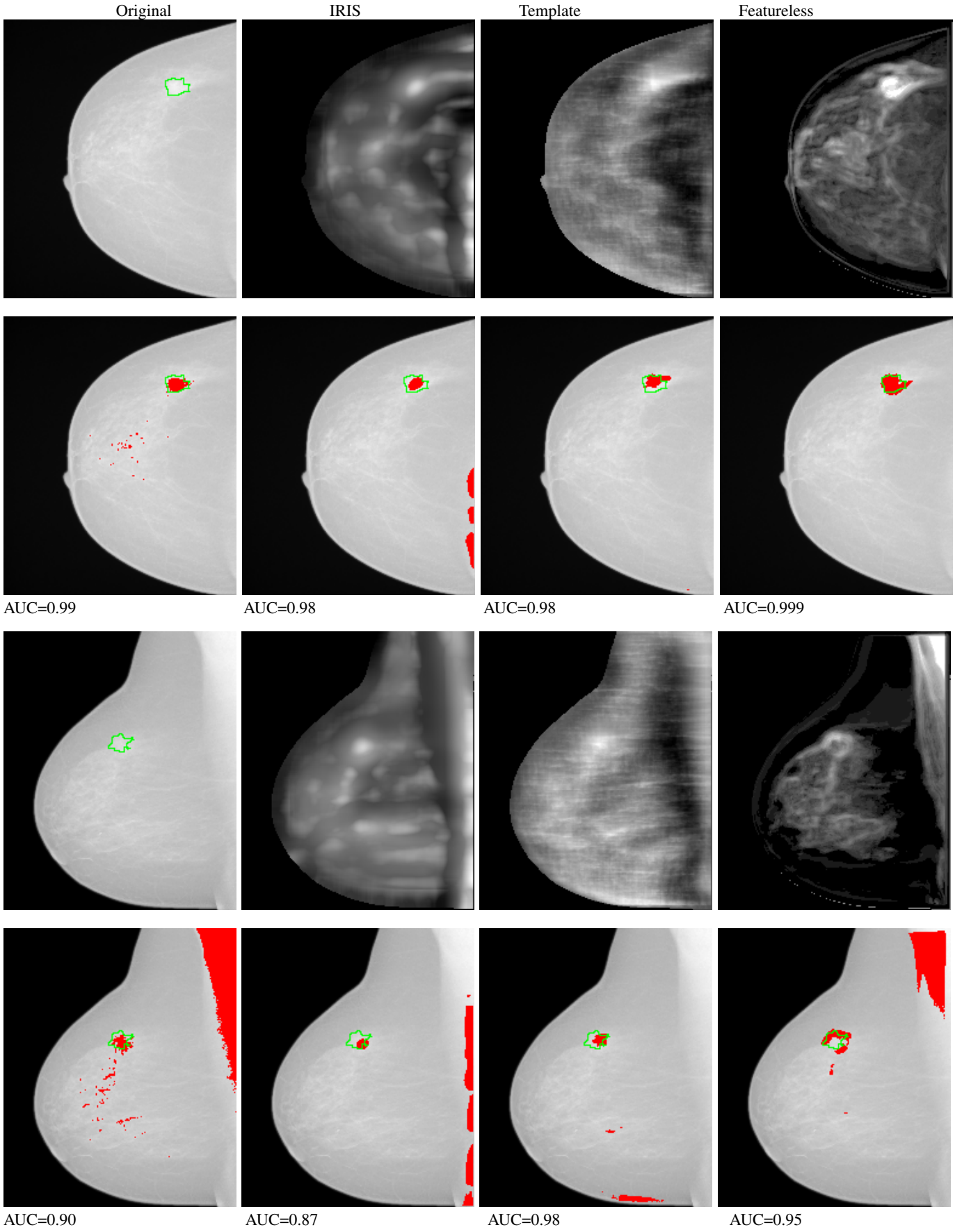


Fig. 7. Upper left is an image with radiologist's marking; the other three images in the same row are the scores given by different three methods. The second row is the detected regions by setting a threshold for the score image. Third and Forth rows are the results of another image. The AUC of each particular method is indicated at the bottom of each row.