

# Abnormalities Detection in Digital Mammography Using Template Matching

Ahmed M. Farrag

Faculty of Computers and Information, Helwan University, Egypt

**Abstract**—Breast cancer affects 1 in 8 women in the United States. Early detection and diagnosis is key to recovery. Computer-Aided Detection (CAD) of breast cancer helps decrease morbidity and mortality rates. In this study we apply Template Matching as a method for breast cancer detection to a novel data set comprised of mammograms annotated according to ground truth. Performance is evaluated in terms of Area Under the Receiver Operator Characteristic Curve (Area Under ROC) and Free-response ROC.

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

## I. INTRODUCTION

Breast cancer is a disease that causes cells of the breast tissue to behave abnormally and grow out of control. They start invading the breast tissue and spreading (metastasizing) to other organs of the body. Breast tumors or Masses can be benign or malignant. Benign masses are characterized by their oval well-defined boundary, while malignant masses have a more speculated boundary.

Breast cancer is the most common cancer among women in the U.S. (excluding skin cancer) [1]. In Egypt, 38.8% of cancer in women is breast cancer (ranking first for women) [2]. Early detection of breast cancer is key to higher survival rates. Women above 45 with average risk of getting the disease are advised to be screened twice every year. However, radiologists' inexperience, fatigue, inattention and haste lead to false-negatives [3].

Computer-Aided Detection (CAD) systems are systems that process digital or digitized images and mark suspicious areas that the radiologist should pay attention to. Breast cancer CAD helps detect otherwise missed breast carcinoma. In a study by [4], the use of CAD resulted in 19.5% more malignant cases being diagnosed. However, [5]'s study showed no change in detection accuracy and recall rates with and without CAD. This can be explained by the difference in performance. "Highly performing" CAD was shown to improve radiologists' performance, while "poor performing" CAD negatively affected their performance [6]. This shows how important designing a highly performing CAD system is.

### A. Data Collection

Our data comprises of only digital mammograms. Our radiologist consultant provides, reviews and annotates each image. She marks the boundaries of the tumor and assigns its category. Markings are done according to ground truth, i.e. all subjects

were biopsied prior to the radiologist's marking. The tumor is categorized according to the "Breast Imaging Reporting and Data System" (BIRADS) scoring system (Table I). A special in-house tool was prepared specifically for this purpose. Both lesions and images were classified by the radiologist according to their BIRAD score.

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.

### B. Related Work

Template Matching is categorized as a model-based method of segmentation, meaning that it involves training and learning. Here, the training is to choose the best method parameters that yield the best result according to some assessment criteria (See section III). As early as 1989, Lai et al. [7] used template matching for detecting circumscribed masses in mammograms. Ng and Bischof in 1992 [8] used the same method. The problem with that approach was the disregard to different masses sizes. This problem is addressed in this paper by multiscale templates described in section II.

The previously mentioned methods use correlation as a similarity measure. A different similarity measure is mutual information. It has been used by Tourassi et al. in 2003 [9].

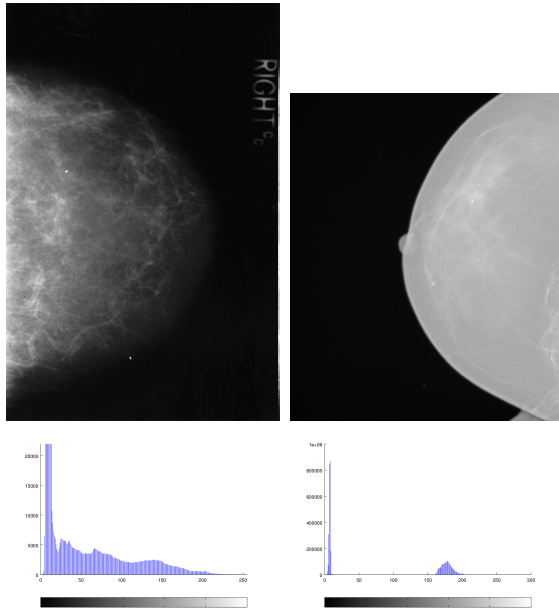
Oliver et al. in 2006[10] and in 2008[11] proposed a probabilistic template to match against. The template is learned from different training tumors.

Brake et al. in 1999[12] proposed a multi-scale approach which we use in our study. Multiple templates of different sizes are used and the correlation to each template is calculated.

The rest of the paper is organized as follows. In section II we detail our approach. In section III, we describe our assessment methods, namely the Area Under the ROC (AUC) and the FROC. In Section IV we show our results. In Section V we conclude the paper and discuss our recommendations for future work.

## II. DESIGN

In this section, we introduce the method used in our study. The method belongs to the pixel-based algorithms. Pixel-based algorithms take as input features the plain grey-level of each pixel and its surrounding pixels up to  $L$  levels. In contrast, region-based algorithms use some image processing techniques to first segment the image into sub-regions which are then used for feature extraction. Classification for pixel-based algorithms is done per pixel, while in region-based algorithms it is done for the entire region.



**Fig. 1:** A figure that illustrates the difference between analogue and digital mammography. To the left is an analogue image and below it is its histogram. The right is the same but with a digital image. It's evident how the gray levels span a much broader range in an analogue image.

Before we start taking grey-level values as our features, preprocessing of all images is carried out. Images are captured using different imaging machines having different sensors with different sensitivity to radiation. Thus, images' grey-level values need to be normalized first so that comparing their values would make sense. In addition to normalization, the breast needs to be segmented away from the background. In our study we used digital mammograms only; those mammograms have no artifacts and the background is almost all zeros making a whole literature on segmenting the breast in analogue mammograms irrelevant. In our study, simple methods like K-means or Otsu's thresholding[13] were sufficient.

### A. Template Matching

Template Matching is carried out by sliding a window of size  $Z \times Z$  containing the desired template over the pixels of the image. Each pixel of the image gets a score according to how similar it and its surrounding pixels are to the template.

This similarity is calculated in terms of correlation, which measures the covariance between the subimage and the template normalized by their variances. Examples of templates that have been used for breast cancer detection in the literature are the spherical template (Figure 2(a))[12] and the hyperbolic secant template (Figure 2(b))[14]

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

$$T(x, y) = \text{sech}(x + y).$$

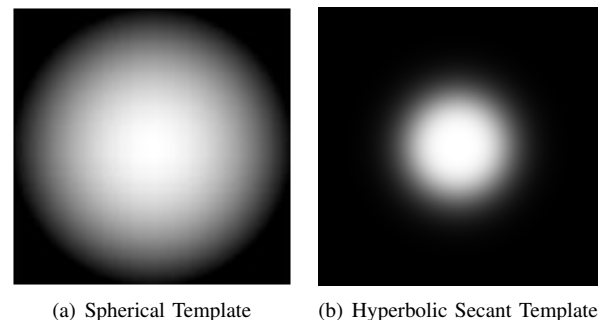
Different template sizes have been tested starting at radius  $R = 4\% \dots 12\%$  of the image height. The multiresolution analysis conducted by Brake et al.[12] has been adopted. The scores for the same pixel produced by correlation of different template sizes have been combined. The final score of the pixel is the maximum of those combined scores. The combination of template sizes (6%, 8%, 10%) yielded best results in terms of an AUC of 0.8656 (AUC is described in III).

## III. ASSESSMENT

### A. AUC

Different classification methods produce scores for every pixel (or region in case of a region-based approach as discussed in II). A certain threshold has to be selected above which the pixel is labeled Class A and below which it is labeled Class B. At different thresholds, some pixels are True Positives (correctly labeled as positives), and some are False Positives. Plotting the True Positive Fraction (TPF) against the False Positive Fraction (FPF) as the decision threshold is varied is called Receiver Operating Characteristic (ROC) curve [15].

The Area Under the ROC Curve (AUC) is a single number that can describe the performance of a classification method. It conveys how separable and distinguishable the two classes have become after using the method. An AUC of one indicates total separability. The AUC of a score image (probability images) is



**Fig. 2:** A rendering of different templates

defined as

$$\widehat{AUC} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(x_i < y_j),$$
$$I_{(c)} = \begin{cases} 1, & \text{if } c \text{ is true} \\ 0, & \text{if } c \text{ is false} \end{cases}$$

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. It should be noted that the AUC assessment is a pixel-based assessment.

### B. FROC

Free Response Operating Characteristic (FROC), on the other hand, assesses the performance in a region-based manner. It is similar to ROC analysis except that in the x-axis there is the number of false positives (regions) per image. Many definitions, thus, exist for what constitutes a detected region. An example of such criterion is a 50% overlap between the detected region and the radiologist's annotation. The common practice in the literature is to assess CADs in terms of the Free-response Receiver Operating Characteristics (FROC) [16].

## IV. RESULTS AND DISCUSSION

We applied the method introduced in Section II to a data set consisting of 50 malignant cases and 100 normals. The assessment method is the average AUC (defined in section III), where the average is taken over images. The average AUC for multiscale template matching with scales 6% 8% and 10% is 0.866.

Figure 4 compares the AUC of the scores of template matching to that of using plain gray-level as a score. The first row shows the probability image(scores image). The second shows the regions detected at a certain threshold.

Figure 5 shows a comparison between template matching and two other in-house developed methods. Different criteria for defining what a detection is exist in the literature. A comparison between 4 of them and our proposed criterion is shown.

One problem with using FROC plots to evaluate pixel-based approaches like template matching is that the resulting curve is non-monotonic. This is caused by the fact that at a given threshold, there are different groups of pixels that merge to for a single marker; but as the threshold is increased, regions would grow and merge with each other until eventually all regions merge into a single big region that occupies the entire breast. This can be seen in Figure 3 by moving from right to left.

## V. CONCLUSION AND FUTURE WORK

In this paper we showed some of the relevant literature. We introduced our novel data set and its collection method. Template Matching with a spherical template has been tested and evaluated using both AUC and FROC as evaluation criteria.

Although Template Matching alone shows no superior results, when it's combined with the scores of other in-house developed methods in an ensemble or a Multiple Classifier System (MCS) the performance was boosted. The MCS we experimented was a simple one. The final score was just a

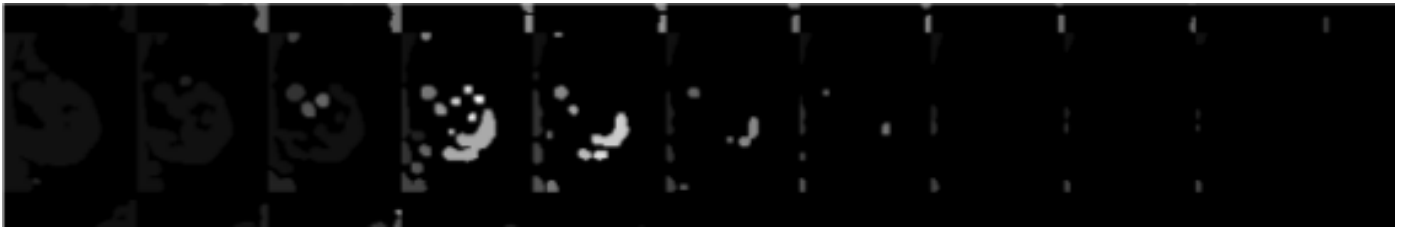
weighted average of Template Matching's score and the two other methods.

For the future we intend to further investigate MCS. In addition, Deep Learning has shown good results with similar problems. We think applying Deep Learning techniques in the literature should yield better performance.

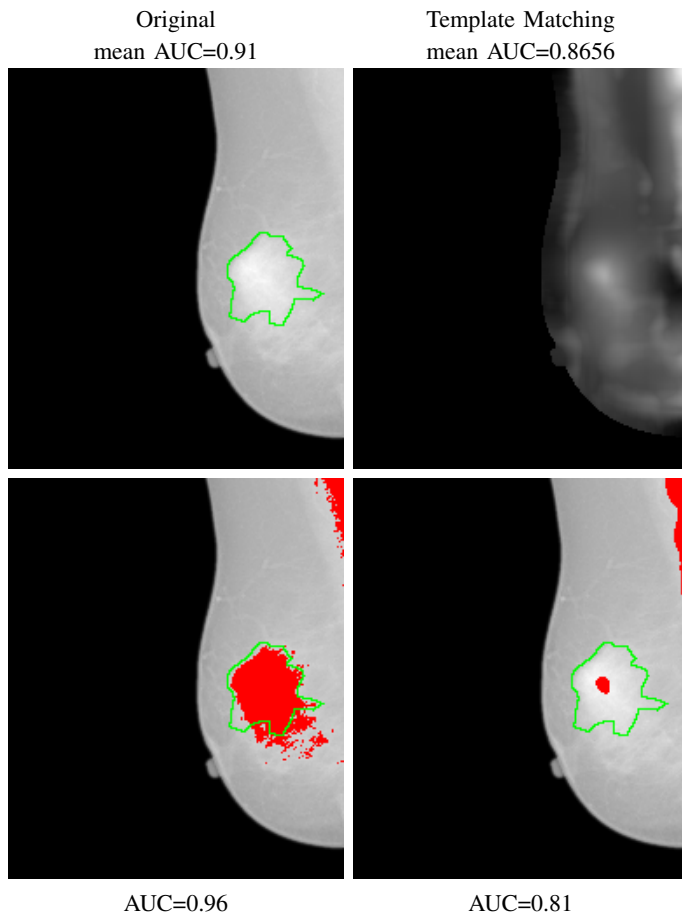
## REFERENCES

- [1] American Cancer Society, *Breast Cancer Facts & Figures 2015-2016*. Atlanta: American Cancer Society, 2015. [Online]. Available: <http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-046381.pdf>
- [2] A. S. Ibrahim, H. M. Khaled, N. N. Mikhail, H. Baraka, and H. Kamel, "Cancer Incidence in Egypt: Results of the National Population-Based Cancer Registry Program," *Journal of Cancer Epidemiology*, vol. 2014, pp. 1–18, 2014. [Online]. Available: <http://www.hindawi.com/journals/jce/2014/437971/>
- [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. 178–194, 2007.
- [4] T. W. Freer and M. J. Ulissey, "Screening Mammography with Computer-Aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center," *Radiology*, vol. 220, no. 3, pp. 781–786, 2001. [Online]. Available: <http://dx.doi.org/10.1148/radiol.2203001282>
- [5] D. Gur, J. H. Sumkin, H. E. Rockette, M. Ganott, C. Hakim, L. Hardesty, W. R. Poller, R. Shah, and L. Wallace, "Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," *J Natl Cancer Inst*, vol. 96, no. 3, pp. 185–190, 2004. [Online]. Available: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve{&}db=PubMed{&}dopt=Citation{&}list{&}uids=14759985>
- [6] B. Zheng, M. Ganott, C. Britton, and C. Hakim, "Soft-Copy Mammographic Readings with Different Computer-assisted Detection Cuing Environments: Preliminary Findings1," *Radiology*, 2001. [Online]. Available: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed{&}cmd=Retrieve{&}dopt=AbstractPlus{&}list{&}uids=1785692547531020097&related=j-h8GByL0PcJ\\$\delimitter"026E30F\\\$papers2://publication/uuid/3D711C9C-BE1F-4858-8FB7-FEF3B5B8222C](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed{&}cmd=Retrieve{&}dopt=AbstractPlus{&}list{&}uids=1785692547531020097&related=j-h8GByL0PcJ$\delimitter)
- [7] S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," *IEEE Transactions on Medical Imaging*, vol. 8, no. 4, pp. 377–386, 1989.
- [8] S. L. Ng and W. F. Bischof, "Automated Detection and Classification of Breast-Tumors," *Computers and Biomedical Research*, vol. 25, no. 3, pp. 218–237, 1992. [Online]. Available: <http://dx.doi.org/10.1109/32.11992HX26400003>
- [9] G. D. Tourassi, R. Vargas-Voracek, D. M. Catarious, and C. E. Floyd, "Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information," *Medical physics*, vol. 30, no. 8, pp. 2123–2130, 2003.
- [10] A. Oliver, J. Freixenet, E. R. E. Denton, R. Zwiggelaar, and C. Science, "Mammographic mass eigendetection," *Medical Image Understanding and Analysis*, pp. 71–75, 2006.
- [11] A. Oliver, J. Freixenet, R. Mart, J. Pont, P. Elsa, E. R. E. Denton, and R. Zwiggelaar, "A Novel Breast Tissue Density Classification Methodology," *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, vol. 12, no. 1, pp. 55–65, 2008.
- [12] G. M. te Brake and N. Karssemeijer, "Single and multiscale detection of masses in digital mammograms," *IEEE transactions on medical imaging*, vol. 18, no. 7, pp. 628–639, 1999.
- [13] N. Otsu, "A threshold selection method from gray-level histograms," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 9, no. 1, pp. 62–66, 1979. [Online]. Available: <http://ieeexplore.ieee.org/xpls/abs{&}all.jsp?{&}amp;arnumber=4310076>
- [14] S. Morrison and L. Linnett, "A Model Based Approach to Object Detection in Digital Mammography," in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, 1999, pp. 182–186 vol.2.
- [15] A. P. Bradley, "The Use Of The Area Under The ROC Curve In The Evaluation Of Machine Learning Algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [16] D. Chakraborty and K. Berbaum, "Observer studies involving detection and localization: Modeling, analysis, and validation," *Medical Physics*, vol. 31, no. 8, pp. 2313–2330, 2004.

[Online]. Available: <http://link.aip.org/getpdf/servlet/GetPDFServlet?filetype=pdf&id=MPHYA6000031000008002313000001>



**Fig. 3:** The detected regions. By moving from right to left (increasing the threshold), different regions grow and merge to form a new bigger region.



**Fig. 4:** Upper left is an image with radiologist's marking; the other image in the same row is the scores given by Template Matching method. The second row is the detected regions by setting a threshold for the score image. The AUC is indicated at the bottom of the row.

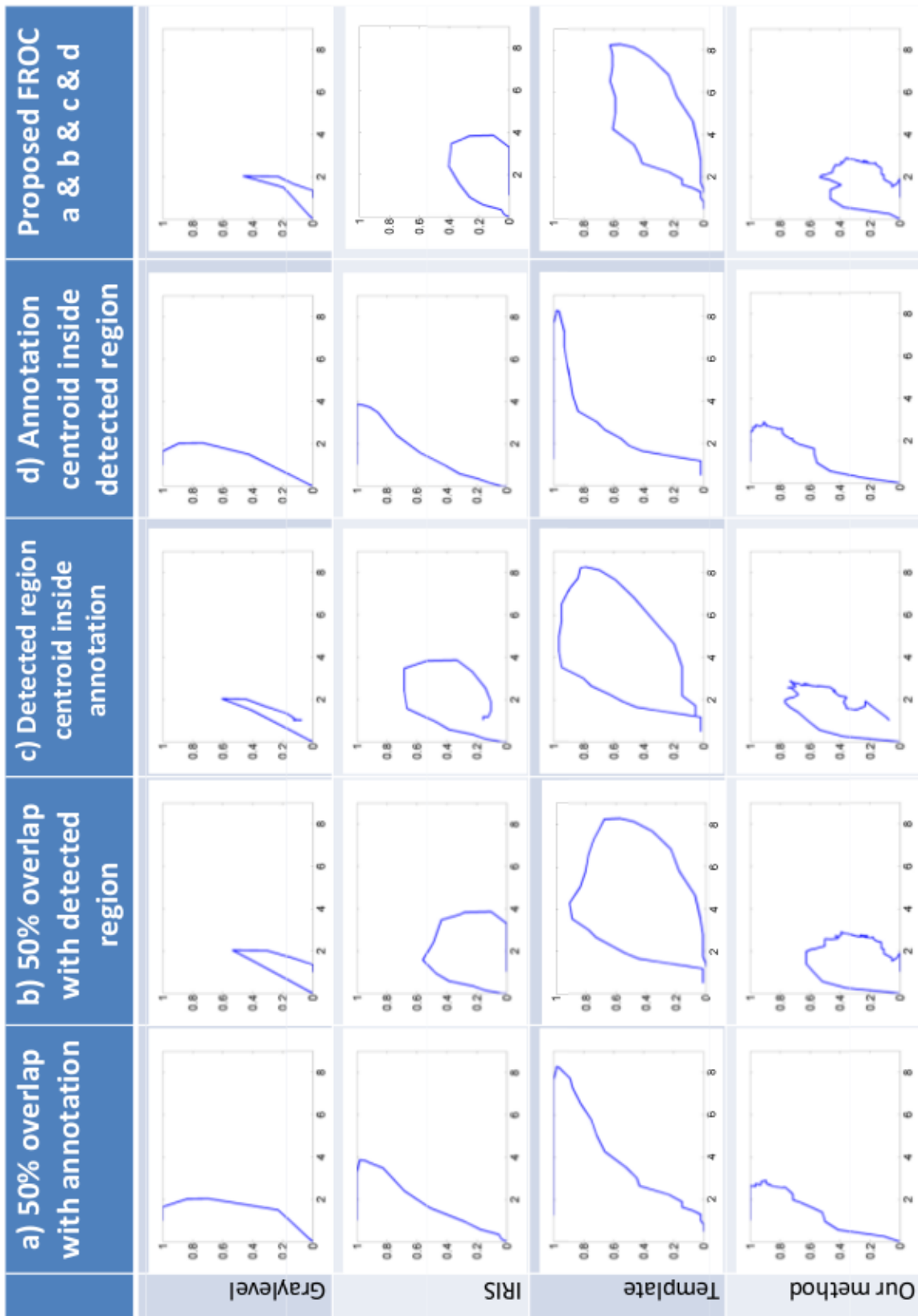


Fig. 5: Conventional FROC results using four criterion for the four detection methods.