POINT PROCESS MODELING FOR DETERMINING DETECTION ACCURACY OF MAMMOGRAPHIC MICROCALCIFICATIONS

María V. Sainz de Cea and *Yongyi Yang*

Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616

ABSTRACT

The occurrence of false positives (FPs) varies greatly from case to case in computerized detection of clustered microcalcifications (MCs) from mammogram images. In this work, we explore using a stochastic modeling approach to estimate the number of individual FPs present in a detected MC lesion. We model the spatial occurrence of FPs in the MC detector output by a Poisson spatial point process, of which the parameters are estimated from the detected objects both in the lesion region and in a reference region in the immediate neighborhood of the lesion. We demonstrate the proposed approach on a set of 188 full-field digital mammography images with two existing MC detectors. The results show that on average the error level in the estimated number of FPs is 3.62 when the actual number of FPs is at the level of 11.38 in a detected MC lesion.

Index Terms— Breast cancer, microcalcifications, false positives, detection accuracy.

1. INTRODUCTION

Microcalcifications (MCs) are tiny calcium deposits that appear as bright spots in mammograms (e.g., Fig. 1(a)). Studies show that clustered MCs can be a sign of non-palpable breast cancer [1], which are reported to be present in 30-50% of the patients diagnosed at this early stage [2]. In spite that MCs are often seen, they are difficult to diagnose accurately. It is reported that only 10% to 40% of the biopsies for MCs are found to be malignant [3]. MCs can be subtle in appearance and can vary greatly in shape and size. There are a number of studies to determine the potential relationship between the features of MCs and the pathology, e.g., [4], [5], [6].

In the literature, there have been great interests in developing computerized methods to aid the diagnosis of MC lesions, which are collectively known as computer-aided diagnosis (CAD). Generally speaking, there are two distinct tasks in a CAD system for MC lesions. One is to determine whether clustered MCs are present or not in a mammogram under consideration (called CADe). The other task is to determine whether a detected MC lesion (either by human or by CADe) is malignant or benign (called CADx).

In practice, the utility of a CAD system is often compromised by the occurrence of false-positives (FPs) in detection. Studies show that there can be many factors contributing to FPs in MC detection [7], e.g. MC-like noise patterns, imaging artifacts, linear structures such as milk ducts, etc. Because of this, there have been great efforts to develop CADe methods for improving the accuracy in MC detection (increasing sensitivity and reducing FP rate), e.g., [8], [9].

Despite these efforts, a major challenge in MC detection is the great inter-subject variability in mammogram characteristics. For example, accurate MC detection is reported to be more difficult in dense breasts [10] or young women [8]. Consequently, the detection accuracy level of individual MCs within a lesion is subject to large case-to-case variations. Conceivably, this variability can directly affect the outcome when the detected MCs in a lesion are subsequently classified by a CADx classifier as malignant or benign. Thus, it is important to assess beforehand how accurate the detections are in a detected MC lesion. When the number of FPs is too high in a detected lesion, the CADx result likely will become unreliable to use in a CADx system.

Traditionally, the detection accuracy of an MC detector is evaluated by the ensemble average over a set of test cases (e.g. using free-response receiver operating characteristic (FROC) curve) [11]. However, to the best of our knowledge, little work exists on how to determine the accuracy of the detections for a specific lesion, which can vary from case to case as noted above. In this work, we develop an estimation approach to assess the accuracy of individual MCs in a detected lesion. We aim to determine how many are FPs among the detected MC objects.

Toward this goal, we exploit the spatial characteristics of the detected objects in a given lesion, and use a Poisson spatial point process to model the random nature of the occurrence of FPs in MC detection. To determine the number of FPs among the detected MCs, we use a maximum likelihood approach to estimate the parameters of the spatial point process based on the detections both in the lesion region and in a reference region in the immediate neighborhood of the lesion. The parameter estimation is done on an image-by-image basis so as to adapt to the noise characteristics of each lesion image. The proposed approach is general and can be applicable to any MC detectors. In the experiments, we demonstrate the proposed approach on the detection results by two existing detectors of which both are well cited in the literature: the difference of Gaussian (DoG) detector in [12] and the support vector machine (SVM) detector in [13]. These two detectors are different in terms

of their detection accuracy, and thus, serve a good test bed for the proposed approach.

2. METHODS

2.1. Motivation

For illustration, we show in Fig. 1(a) an example of a mammogram region with clustered MCs; the corresponding detection output of this region by an MC detector (SVM) [13] is shown in Fig. 1(b). While the MCs are notably enhanced in the detector output, there are also multiple bright objects which do not correspond to actual MCs. Depending on the operating threshold used, some of these latter objects could be falsely detected as MCs. As noted in the introduction, both the MCs and FPs can vary greatly from case to case even when the same detector is used. Our objective is to determine the number of FPs that may be present among the detected objects in a lesion.

2.2. Problem formulation

Consider a lesion region A, which contains clustered MCs in a mammogram image. Assume that, for locating the individual MCs, an MC detector $f(\cdot)$ has been applied to the image region such that a set of detected objects (i.e., potentially MCs) is obtained in A. We want to estimate the number of true MCs among the set of detections. For convenience, let's denote each detected object by index i, $i = 1,\ldots,n$ where n is the number of detected objects, and let $y^{(i)}$ denote its unknown label (1 for being a true MC, and 0 otherwise). Then, the number of true MCs among all the detections is given by:

$$M = \sum_{i=1}^{n} y^{(i)} \tag{1}$$

Our goal is to determine the value of M.

Note that the number of false positives (FPs) among the detections is given by $n - M$. Thus, from this point on we refer to the estimation of the number of true positives (TPs) indistinctly from that of FPs.

2.3. Spatial point process model

Consider a set of n detected objects in lesion region A. We aim to estimate the number of TPs (i.e., M) by modeling the occurrence of FPs in A as a spatial point process (SPP), which is a stochastic model for describing the random distribution pattern of a set of points in a d-dimensional space ($d = 2$ in our application) [14].

Specifically, we model the spatial distribution of the FPs within A, (neither their locations nor their number of occurrences are known), by a Poisson point process [15]. Such a process is characterized by the following two properties: i) for a bounded spatial region B, the number of points contained within B follows a Poisson distribution of which the mean parameter is proportional to the area of B; and ii) for two disjoint regions B_1 and B_2, the number of points within B_1 is independent of that within B_2.

To determine the number of FPs in lesion region A, a challenge is that both FPs and TPs can occur simultaneously within A, but neither of the two is known. To resolve this difficulty, we introduce a reference region R, which is located in the immediate vicinity of A from the same mammogram, as illustrated in Fig. 2. This reference region is assumed to have the following properties: i) it does not contain any true MCs; hence, only FPs will be detected by the MC detector in R; ii) the image noise characteristics in R are similar to those in lesion region A; thus, the occurrence of FPs follows a common random process in the two regions. To quantify the random process of FPs, we apply the MC detector to both A and R. Let N_r denote the number of resulting detections in R. Then N_r obeys the following Poisson distribution:

$$p(N_r \mid \lambda) = \frac{(\lambda A_r)^{N_r}}{N_r!} e^{-(\lambda A_r)} \tag{2}$$

where A_r denotes the area of R, and λ denotes the rate parameter which is the expected number of FPs per unit area.

On the other hand, for lesion region A, let M be the number of detected MCs as in (1), and N_l the total number of detections. Then $N_l - M$ is the number of FPs, and N_l obeys the following distribution:

$$p(N_l \mid \lambda, M) = \frac{(\lambda A_l)^{N_l - M}}{(N_l - M)!} e^{-(\lambda A_l)} \tag{3}$$

where A_l denotes the area of lesion region A.

Given that the two regions A and R are non-overlapping, the joint distribution of the detections in the two regions can be written as:

$$p(N_r, N_l \mid \lambda, M) = \frac{\lambda^{N_r + N_l - M} A_r^{N_r} A_l^{N_l - M}}{N_r!(N_l - M)!} e^{-\lambda(A_r + A_l)} \tag{4}$$

Our goal then becomes to estimate the parameters λ (rate of FPs) and M (the number of TPs) in (4). Given the observations N_r and N_l, we use maximum likelihood (ML) estimation [16] to determine λ and M. That is,

$$[\hat{\lambda}, \hat{M}] = \arg\max_{\lambda, M} \log p(N_r, N_l \mid \lambda, M) \tag{5}$$

In our experiment, we used the interior-point algorithm [17] for the optimization problem in (5)

3. EXPERIMENTS AND RESULTS

3.1. Performance evaluation

To evaluate the estimation performance, for a given lesion with n detected objects, we compare the estimated number of TPs (or FPs) among these n detections against the actual number. For quantifying the estimation accuracy, we use the following relative error:

$$e_P = \frac{\left| M - \hat{M} \right|}{n} \tag{6}$$

where M is the actual number of TPs and $\hat{M}$ is the estimated number.

Note that the fraction of TPs (TPF) among the n detections is given by M/n. Thus, the relative error e_P simply corresponds to the difference between the actual and the estimated TPF in the detection results of a given lesion. Furthermore, despite that the relative error e_P is defined in terms of the number of TPs (i.e., M) in (6), it can be shown that the relative error e_P also corresponds to the difference between the actual fraction of FPs (FPF) and the estimated FPF among the detected objects. To summarize the estimation accuracy, we report the mean and median values of e_P over all the lesions in the dataset.

3.2 MC detectors for demonstration

To demonstrate the proposed approach for determining the accuracy in MC detection, we consider two existing MC detectors, both of which have been well cited in the literature. The first is an SVM detector developed in [13], which is based on supervised learning. The second is the DoG detector [12], which is of low computational complexity for MC detection.

For MC detection in a lesion image, the output of the detector in use (SVM or DoG) was compared against an operating threshold $T = \mu + c\sigma$, where μ and σ are the mean and standard deviation determined from the detector output of the image. The parameter c is a constant value for setting the operating threshold. To assess the performance of the proposed approach at different FP levels, the operating threshold T was adjusted so that the sensitivity level in detection was varied. Afterward, the proposed approach was applied to estimate the detection results by the detectors.

3.4 Mammogram dataset

For this study we make use of a set of full-field digital mammography (FFDM) images collected by the Department of Radiology at the University of Chicago. It consisted of 188 images from 95 cases (43 malignant, 52 benign), all containing clustered MCs. These images are 100μm/pixel in resolution. Most of the cases had both craniocaudal and mediolateral oblique views. The MCs in each mammogram were manually identified by a researcher with more than 15 years of experience in mammography research and with special training on interpreting mammograms. In total there were 8,979 MCs marked. For the purpose of evaluating the accuracy in the detected MCs, the lesion regions of clustered MCs in these mammograms were marked out by a bounding circle with a diameter of 1cm, 2cm, or 3cm (according to the lesion size) so that all the marked MCs were contained inside the circle; for those elongated lesions, an ellipse of equal area was used in place of the bounding circle. A reference region is needed for each lesion. In our experiment, this region was set to be the immediate annulus region outlier of the lesion of which the area is equivalent to that of a disk of 3cm in diameter. For illustration, an example is shown in Fig. 2 for a lesion and its reference region.

3.4 Results

We summarize in Table 1 the estimation results obtained by the proposed approach when the SVM detector was used for MC detection on the lesions in the dataset. For demonstration, the estimation results were given for four different operating points with detection sensitivity ranging from 60% to 75%. It can be seen that, as the sensitivity level varied from 60% to 75%, the FPF level in the detection results is increased from 0.1735 to as high as 0.4260. However, the estimation error e_P obtained by the spatial point process maintains a rather consistent level within this range. In particular, with FPF at 0.4260, the mean and median of e_P are 0.1622 and 0.1429, respectively.

Similarly, we summarize in Table 2 the estimation results when the DoG detector was used for MC detection. As can be seen, with the sensitivity level varied from 60% to 75%, the FPF level is increased from 0.2005 to 0.5094. Within this operating range, the mean error e_P achieved by the proposed estimation approach is between 0.1474 and 0.1741. Moreover, it is noted that in both Tables 1 and 2 the median value of e_P is smaller than its corresponding mean value at each of the four sensitivity levels. This indicates that the error distribution is skewed toward the left (i.e., lesions having smaller estimation errors). Note that at sensitivity level 75% the FPF level is already above 50% for the DoG detector. For practical purposes further higher sensitivity levels are not considered in this study.

In Fig. 3 we show a scatter-plot of the estimation results on the individual lesions obtained by the proposed approach when the SVM detector was used for MC detection with sensitivity level set at 70%. In this plot, each point corresponds to a lesion, of which the y coordinate is the estimated number of TPs ($\hat{M}$) and the x coordinate is the actual number of TPs (M). Thus, a point on the 45° line represents a perfect match between M and $\hat{M}$. For the plot a logarithmic scale was used in order to accommodate for the large range in the number of MCs among different lesions. The correlation coefficient between M and $\hat{M}$ is 0.9467. A paired t-test comparison reveals no statistically significant difference between the mean values of M and $\hat{M}$ (p-value = 0.8523), indicating that there is no systematic bias in the estimate.

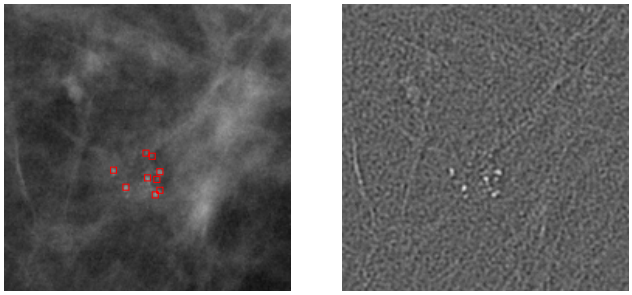
From Fig. 3 it can be seen that the estimation error varies with the number of TPs in a lesion. In particular, it tends to be larger for the lesions with fewer TPs. This is more related to the number of detections in the cluster. To further examine this, we divided the dataset into three groups: group one having lesion with 10 or fewer detections (76 lesions), group two having lesions between 11 and 30 detections (89 lesions), and group three having lesions with 31 or more detections (35 lesions). We then calculated the estimation error e_P for the lesions in each of the three

groups. In Fig. 4, we show a box-car plot of the distribution of e_P obtained in each group. As can be seen, the error decreased from group 1 to group 3; for example, the median value of e_P was 0.1667 in group 1, compared to 0.1053 in group 3.

We believe that the above observation can be explained as follows: in the proposed estimation approach, we need to determine the parameters of the underlying SPP model based on the number of detected objects in a lesion area. Among the three groups considered above, the number of detections increased from group 1 to group 3, and so did the number of FPs, with the average being 2.13 in group 1, 6.5 in group 2, and 17.65 in group 3. This in turn helped improve the estimation accuracy of the model parameters as the number of FPs increased.

4. CONCLUSIONS

We proposed a spatial point process approach for estimating the number of FPs detected in an MC lesion. This approach was demonstrated to be effective on two different MC detectors over a range of sensitivity levels. The mean and median of the achieved estimation error were within 17% and 15%, respectively, of the number of detections in a lesion for both detectors. These results indicate that the proposed approach can be feasible for assessing the accuracy of detected MCs within a lesion in a CADe framework.



(a) (b)

Fig. 1 (a) A mammogram ROI containing clustered MCs (marked by red squares), and (b) its detection output by an MC detector.

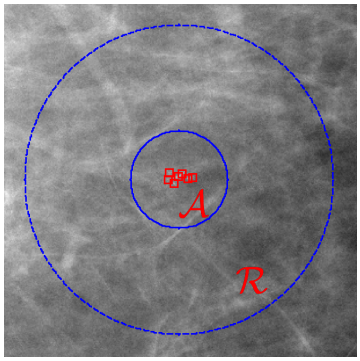


Fig. 2 Example of lesion region (A) and its reference region (R). The MCs are marked by red squares.

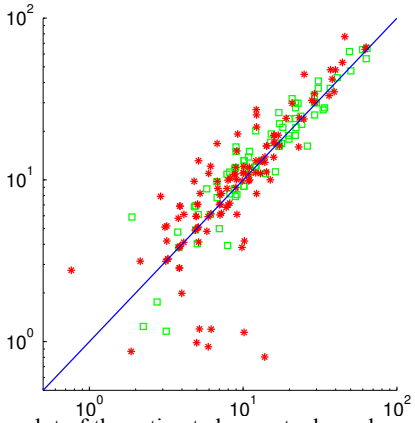


Fig. 3 Scatter plot of the estimated vs. actual number of TPs (red * symbols: malignant cases; green squares: benign cases).

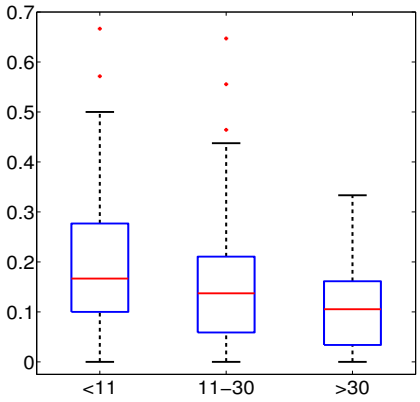


Fig. 4 Box-car plots of the estimation error e_P according to the number of detections in each lesion in the dataset. Group 1 lesions have 10 or fewer detections, group 2 lesions have detections between 11 and 30, and group 3 lesions have more than 30 detections.

Table 1 Estimation error e_P in SVM detection results

Sensitivity	Mean FPF	Error e_P	
0.6	0.1735	Mean Median	0.1591 0.1176
0.65	0.2316	Mean Median	0.1627 0.1396
0.7	0.3314	Mean Median	0.1641 0.1429
0.75	0.4260	Mean Median	0.1622 0.1429

Table 2 Estimation error e_P in DoG detection results

Sensitivity	Mean FPF	Error e_P	
0.6	0.2005	Mean Median	0.1474 0.0952
0.65	0.2779	Mean Median	0.1741 0.1250
0.7	0.3937	Mean Median	0.1724 0.1539
0.75	0.5094	Mean Median	0.1697 0.1437

5. REFERENCES

[1] E. A. Sickles, "Mammographic features of" early" breast cancer," *American Journal of Roentgenology,* vol. 143, no. 3, pp. 461-464, 1984.

[2] H.-P. Chan, K. Doi, C. J. Vyborny, K.-L. Lam, and R. A. Schmidt, "Computer-Aided Detection of Microcalcifications in Mammograms Methodology and Preliminary Clinical Study," *Investigative Radiology,* vol. 23, no. 9, pp. 664-671, 1988.

[3] E. Fondrinier, G. Lorimier, V. Guerin-Boblet, A.-F. Bertrand, C. Mayras, and N. Dauver, "Breast Microcalcifications: Multivariate Analysis of Radiologicand Clinical Factors for Carcinoma," *World journal of surgery,* vol. 26, no. 3, pp. 290-296, 2002.

[4] U. Kettritz, G. Morack, and T. Decker, "Stereotactic vacuum-assisted breast biopsies in 500 women with microcalcifications: radiological and pathological correlations," *European journal of radiology,* vol. 55, no. 2, pp. 270-276, 2005.

[5] M. V. Karamouzis *et al.*, "Non−palpable breast carcinomas: Correlation of mammographically detected malignant−appearing microcalcifications and molecular prognostic factors," *International journal of cancer,* vol. 102, no. 1, pp. 86-90, 2002.

[6] R. Nakayama, Y. Uchiyama, R. Watanabe, S. Katsuragawa, K. Namba, and K. Doi, "Computer-aided diagnosis scheme for histological classification of clustered microcalcifications on magnification mammograms," *Medical physics,* vol. 31, no. 4, pp. 789-799, 2004.

[7] T. Ema, K. Doi, R. M. Nishikawa, Y. Jiang, and J. Papaioannou, "Image feature analysis and computer−aided diagnosis in mammography: Reduction of false−positive clustered microcalcifications using local edge−gradient analysis," *Medical Physics,* vol. 22, no. 2, pp. 161-169, 1995.

[8] H.-D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey," *Pattern recognition,* vol. 36, no. 12, pp. 2967-2991, 2003.

[9] R. M. Rangayyan, F. J. Ayres, and J. E. L. Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," *Journal of the Franklin Institute,* vol. 344, no. 3, pp. 312-348, 2007.

[10] W. M. Morrow, R. B. Paranjape, R. M. Rangayyan, and J. E. L. Desautels, "Region-based contrast enhancement of mammograms," *IEEE transactions on Medical Imaging,* vol. 11, no. 3, pp. 392-406, 1992.

[11] H. Bornefalk and A. B. Hermansson, "On the comparison of FROC curves in mammography CAD systems," *Medical physics,* vol. 32, no. 2, pp. 412-417, 2005.

[12] J. Dengler, S. Behrens, and J. F. Desaga, "Segmentation of microcalcifications in mammograms," *IEEE Transactions on Medical Imaging,* vol. 12, no. 4, pp. 634-642, 1993.

[13] 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 Transactions on Medical Imaging,* vol. 21, no. 12, pp. 1552-1563, 2002.

[14] J. Møller and R. P. Waagepetersen, "Modern statistics for spatial point processes," *Scandinavian Journal of Statistics,* vol. 34, no. 4, pp. 643-684, 2007.

[15] A. Baddeley, I. Bárány, and R. Schneider, "Spatial point processes and their applications," *Stochastic Geometry: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 13–18, 2004,* pp. 1-75, 2007.

[16] F. W. Scholz, "Maximum Likelihood Estimation," in *Encyclopedia of Statistical Sciences*: John Wiley & Sons, Inc., 2004.

[17] S. J. Wright, *Primal-dual interior-point methods*. Siam, 1997.

[18] B. Efron and G. Gong, "A leisurely look at the bootstrap, the jackknife, and cross-validation," *The American Statistician,* vol. 37, no. 1, pp. 36-48, 1983.