Published in final edited form as:

Phys Med Biol. ; 63(4): 045014. doi:10.1088/1361-6560/aaa4c.

Locally Adaptive Decision in Detection of Clustered Microcalcifications in Mammograms

María V. Sainz de Cea¹, Robert M. Nishikawa², and Yongyi Yang^{1,*}

¹Medical Imaging Research Center, Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616, USA

²Department of Radiology, University of Pittsburgh, Pittsburgh, PA 15231, USA

Abstract

In computer-aided detection or diagnosis of clustered microcalcifications (MCs) in mammograms, the performance often suffers from not only the presence of false positives (FPs) among the detected individual MCs but also large variability in detection accuracy among different cases. To address this issue, we investigate a locally adaptive decision scheme in MC detection by exploiting the noise characteristics in a lesion area. Instead of developing a new MC detector, we propose a decision scheme on how to best decide whether a detected object is an MC or not in the detector output. We formulate the individual MCs as statistical outliers compared to the many noisy detections in a lesion area so as to account for the local image characteristics. To identify the MCs, we first consider a parametric method for outlier detection, the Mahalanobis distance detector, which is based on a multi-dimensional Gaussian distribution on the noisy detections. We also consider a non-parametric method which is based on a stochastic neighbor graph model of the detected objects. We demonstrated the proposed decision approach with two existing MC detectors on a set of 188 full-field digital mammograms (95 cases). The results, evaluated using free response operating characteristic (FROC) analysis, showed a significant improvement in detection accuracy by the proposed outlier decision approach over traditional thresholding (the partial area under the FROC curve increased from 3.95 to 4.25, p -value $< 10^{-4}$). There was also a reduction in case-to-case variability in detected FPs at a given sensitivity level. The proposed adaptive decision approach could not only reduce the number of FPs in detected MCs but also improve case-to-case consistency in detection.

Keywords

Computer-aided diagnosis (CAD); clustered microcalcifications (MCs); adaptive decision; outlier detection; mammography

1. Introduction

Apart from skin cancer, breast cancer is the most common cancer diagnosed among US women, accounting for 30% of cancer cases. It is also the second leading cause of cancer

Asterisk indicates corresponding author. yangyo@iit.edu.

death among women after lung cancer (*American Cancer Society. Cancer Facts and Figures. Atlanta, GA, 2017, n.d.*), accounting for 14% of the cancer deaths. Screening mammography, which involves X-ray imaging of the breast, is currently one of the most widely used tools for identification of suspicious abnormalities for follow-up and subsequent diagnosis (*American Cancer Society. Cancer Facts and Figures. Atlanta, GA, 2017, n.d.*).

Clustered microcalcifications (MCs) can be a major sign of non-palpable breast cancer. Approximately 50% of the cases diagnosed at this early stage are detected exclusively by the presence of MCs, revealing up to 90% of ductal carcinoma *in situ* (Scimeca et al., 2014). For this reason, MCs are one of the most important diagnostic markers of breast lesions (Naseem et al., 2015). MCs are tiny calcium deposits that appear as bright spots in mammograms (Fig. 1). While being very small, individual MCs can vary greatly in shape, size, and have low contrast relative to the surrounding tissue (Cheng et al., 2003). As a result, accurate detection of MCs has been a challenging task in computer-aided detection (CADE) and diagnosis (CADx) of MC lesions. Because of this, great efforts have been made in developing automated computer algorithms for MC detection (Cheng et al., 2003; Rangayyan et al., 2007; Gur et al., 2004; Elter and Horsch, 2009; Tang et al., 2009; Miranda and Felipe, 2015; Apostolopoulos et al., 2014; Pereira et al., 2014; Bria et al., 2014). Depending on the methodology used, these detection algorithms can range from image enhancement (e.g. McLoughlin et al., 2004; Linguraru et al., 2006), stochastic modeling (e.g. Jing et al., 2011; Yu et al., 2006), to machine learning (e.g. El-Naqa et al., 2002; Wang et al., 2016; Oliver et al., 2012).

Despite these advances in detection algorithms, several challenges still remain in MC detection in practice. One such challenge is the presence of many false positives (FPs) when operated at a reasonable sensitivity level. Studies show that there can be many causes for FPs in MC detection, which include imaging artifacts, or anatomical structures such as fibrous strands, breast borders, or hypertrophied lobules that look like MCs, etc (Mordang et al., 2016; Cheng et al., 2003). Another challenge is the great inter-patient variability in mammogram characteristics. For example, MC detection is noted to be more difficult in denser breasts (Jalalian et al., 2013); in Karahaliou et al. (2014), it is reported that the presence of dense breast parenchyma results in low specificity values and thus unnecessary biopsies.

In the literature there exist a number of studies that aim to reduce the occurrence of FPs in MC detection. Most of these studies focus on improving the accuracy of detected MC clusters in a mammogram (Mordang et al., 2016; Wang et al., 2013, 2016; Bria et al., 2014; Sainz de Cea and Yang, 2015). While CADE algorithms can achieve high sensitivity in detecting MC clusters, identifying individual MCs has proven to be far less accurate due to the great variability in the imaging characteristics among different cases. In our recent work (Sainz de Cea et al., 2017), it was demonstrated that both the fraction of FPs and the sensitivity in detected individual MCs in a set of MC lesions can largely vary among different patients at a common operating point of the MC detector.

In this work, we aim to address the issue of variability in detection of individual MCs by developing a locally adaptive decision approach in the MC detector output. This is important

when one wants to further classify a detected MC cluster as belonging to a benign or a malignant process, because the accuracy of detecting individual MCs can affect the classification accuracy of the cluster (Sainz de Cea et al., 2017). In a CADx system, the detected MCs in a cluster are subsequently further analyzed (by a pattern classifier) for being associated with a benign or malignant lesion. Thus, accurate detection of individual MCs is critical to achieving consistent performance in the presence of patient variability by the CADx system.

In MC detection, the conventional approach is first to apply an MC detector (e.g. a machine learning or image enhancement detector) to a mammogram image under consideration. Afterward, the detector output $f(\cdot)$ is compared against a prescribed threshold T to determine whether an MC is present or not at each location in the image. This operating threshold T is typically set to achieve a target sensitivity level (over an ensemble of cases). However, as noted above, due to the variability associated with both noise characteristics and MC signals among different cases, the resulting detection accuracy from a common threshold T will vary considerably among different mammogram images. To deal with this issue, in our proposed decision scheme we adopt a statistical outlier detection approach for determining the MC objects in the detector output. Unlike the conventional approach of uniform thresholding, this decision scheme directly takes into account the noise characteristics in the detector output for a given image, thereby achieving locally adaptive decision in the detection.

We note that the purpose of this work is not to develop a new MC detector. Instead, we aim to develop a new decision scheme for identifying the detected MC objects in a given detector output and rejecting false detections, which are due mainly to noise. That is, this technique is intended to be applicable to existing MC detectors for dealing with the challenge of variability in detecting individual MCs in a lesion area. We first explored this decision scheme in our preliminary work in Sainz de Cea and Yang (2016), where we used a stochastic neighbor graph (SNG) (Janssens, 2013) for outlier detection. In this work, we further develop this approach along with a parametric outlier detection approach for our locally adaptive decision scheme in MC detection. We also fully evaluate this decision scheme by examining both the overall detection accuracy and the case variability in detection performance.

It is noted that outlier detection methods were previously applied in development of detectors for MC detection in mammograms. In Gurcan et al. (1997), higher order statistics were used for identifying MC clusters. In Thangavel and Mohideen (2010), semi-supervised k-means clustering was used to detect outlier clusters using shape features, where outlier removal was used to improve the classification performance of the CAD system. In Gunawan (2001), a quartile outlier method was used to detect outliers in the wavelet domain. The presented work is different from these studies in that we are developing a decision strategy for improving the accuracy of detected individual MCs in the detector output of a lesion area. Outlier detection has also been used in several medical image segmentation applications. For example, outlier detection was applied in Prastawa et al. (2004) for automatic brain tumor segmentation from MR images where the Mahalanobis distance was used for detecting multidimensional outliers; in Van Leemput et al. (2001), a robust W-

estimator for outlier detection was used for segmenting multiple sclerosis (MS) lesions from MR images; in Schmidt et al. (2012), an iterative outlier detection approach was developed for detecting MS lesions where a distribution was assumed for both normal and outlier classes. Outlier detection has also been used for object detection. For example, in Desclée et al. (2006) a chi-square approach for outlier selection was used to detect forest changes; in Zhou et al. (2013), outlier detection was applied for detecting moving objects.

The rest of the paper is organized as follows: in Sect. 2, we describe the formulation of the proposed locally adaptive decision approach and present the outlier detection methods used; in Sect. 3, we describe the evaluation procedure used for the proposed approach. The evaluation results are presented in Sect. 4. Finally, a discussion is given in Sect. 5 and the conclusion is given in Sect. 6.

2. Methodology of Locally Adaptive Detection Decision

2.1. Motivation and problem formulation

To motivate the problem, we show in Fig. 1 the detector output (SVM detector in El-Naqa et al. (2002)) of two example ROIs from different patients containing clustered MCs. As can be seen, while some of the individual MCs are hardly visible in the mammogram ROIs, they are notably enhanced in the detector output. However, besides the MCs, there are also numerous bright (noisy) spots present in the detector output. The noise patterns are also noted to be different in the two example ROIs. Depending on the operating threshold T used for the detection, some of these spots could be falsely detected as MCs (i.e., FPs) while some of the MCs could be missed by the detector (as to be seen later in Fig. 5 in the results). Moreover, due to the difference in both noise characteristics and MC signals, the detection accuracy of individual MCs can vary in these two clusters.

In this work, we aim to address this variability issue in MC detection by adopting a locally adaptive detection approach in the detector output. As observed from the detector output in Fig. 1, the MCs in a cluster can even differ among themselves in terms of their shape and size. Nevertheless, the MCs tend to be brighter and larger in extent than most of the noisy spots (potential FPs) in the image background. Our decision scheme is to identify automatically the MCs by exploiting their differences from the noisy spots in the detector output in a lesion area.

Specifically, the proposed decision scheme works as follows: In order to exploit the noise characteristics in a given image, we first locate from the detector output a set of noisy spots, denoted by S , which could potentially be detected as FPs. Note that the MCs in the image, not known *a priori*, are likely also detected and contained in set S . Our goal then is to separate the MCs from the rest of the detections in S .

To ensure a high detection sensitivity (i.e., minimizing missed MCs), we need to keep the operating threshold T very low in the detector output (so that all the MCs are contained in S). However, this will unduly increase the number of (noisy) detections in S . Consequently, given that the number of MCs is typically limited in a cluster, the detections in S are overwhelmingly dominated by noisy spots. In our decision scheme, we treat the MCs as

statistical outliers (as noted above to be brighter and larger than noisy spots) among the detections in S , and apply outlier detection algorithms to identify the MCs.

In the literature there have been a wide variety of techniques developed for detecting statistical outliers from a set of data samples (Radovanović et al., 2015; Branch et al., 2013; Jiang et al., 2001; Lukashevich et al., 2009; Ramaswamy et al., 2000). These techniques range from parametric methods to non-parametric methods. In a parametric method, an explicit probability distribution is typically used to characterize the (nonoutlier) data samples (Marsland, 2001; Filzmoser, 2016; Filzmoser et al., 2005), whereas no explicit prior distribution is needed in a non-parametric method (Rousseeuw and Leroy, 2005; Cao et al., 2014).

In this work, we investigate using two outlier detection methods, one parametric and one non-parametric, for our adaptive decision scheme in MC detection. In the first method, we use a multi-dimensional Gaussian distribution to model the detected objects in the detector output. In the second method, we use a stochastic neighbor graph (SNG) to model the similarities among the data samples. Note that these outlier detection methods extract the necessary statistical parameters automatically from the set of data samples under consideration. Thus, no additional training involving other cases is needed. We explain these methods in detail next.

For convenience, in subsequent development we denote the detected objects in S by a set of data points as $\{\mathbf{x}_i, i = 1, \dots, N\}$ in a k -dimensional vector space. The specific definition of these data points is to be given in detail subsequently in Section 2.4.2.

2.2. Mahalanobis distance outlier detector

The Mahalanobis distance (Mahalanobis, 1936) is a measure of the distance of an observed data point from a multidimensional Gaussian distribution $\mathcal{N}(\mu, \Sigma)$, where μ and Σ denote the mean vector and covariance matrix of the distribution, respectively. Consider data sample $\mathbf{x} \in \mathbb{R}^k$. Its Mahalanobis distance from the distribution is defined as

$$\mathcal{D}(\mathbf{x}) = \sqrt{(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)} \quad (1)$$

The variable $\mathcal{D}^2(\mathbf{x})$ is known to follow a chi-square distribution with k degrees of freedom for inlier samples (Lancaster and Seneta, 2005). This property has been widely used for detection of outliers in high dimensional spaces (Todeschini et al., 2013; Zhang et al., 2016).

To apply the Mahalanobis distance detector to our problem, we first need to estimate the parameters μ and Σ from the data samples $\{\mathbf{x}_i, i = 1, \dots, N\}$. A straightforward approach for this would be to use maximum likelihood (ML) estimation (Scholz, 1985). However, the presence of outliers among the data samples is known to bias the ML estimate (Glenn and Zhao, 2007). Out of this consideration, we employ instead a widely used local ML estimation approach (Tibshirani and Hastie, 1987; Eguchi and Copas, 1998; Schubert et al., 2014) according to which the parameters are obtained by maximizing the following modified log-likelihood function

$$L(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \mu, \Sigma) = \sum_{i=1}^N K(\mathbf{x}_i, \mathbf{t}) \log(f(\mathbf{x}_i; \mu, \Sigma)) \quad (2)$$

where $K(\mathbf{x}_i, \mathbf{t})$ is a kernel function centered at $\mathbf{t}$, and $f(\mathbf{x}_i; \mu, \Sigma)$ denote the pdf of the distribution $\mathcal{N}(\mu, \Sigma)$.

In (2), the kernel function is used to assign a higher weight to those data samples that are closer to $\mathbf{t}$ than those further away. The estimates of the parameters μ and Σ are then given by:

$$\hat{\mu} = \sum_{i=1}^N w_i \mathbf{x}_i, \quad \hat{\Sigma} = \sum_{i=1}^N w_i (\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^T \quad (3)$$

where

$$w_i = \frac{K(\mathbf{x}_i, \mathbf{t})}{\sum_{j=1}^N K(\mathbf{x}_j, \mathbf{t})} \quad (4)$$

In this work, we use the following multivariate Gaussian kernel as in (Scott and Sain, 2005; Hyndman et al., 2004):

$$K(\mathbf{x}_i, \mathbf{t}) = \exp\left[-\frac{1}{2}(\mathbf{x}_i - \mathbf{t})^T \Gamma^{-1}(\mathbf{x}_i - \mathbf{t})\right] \quad (5)$$

where $\Gamma = \text{diag}(\gamma_1, \gamma_2, \dots, \gamma_k)$. The diagonal elements are set as (Terrell and Scott, 1992; Bowman and Azzalini, 1997):

$$\gamma_i = \mathcal{S}_i \left[\frac{4}{(k+1)n} \right]^{1/(k+4)} \quad (6)$$

for $i = 1, \dots, k$, where $\mathcal{S}_i$ is the sample standard-deviation of the i -th component of the data vector $\mathbf{x}$. Furthermore, for the vector $\mathbf{t}$ in the kernel function, each of its components is set to be the median value of the corresponding component of the samples $\mathbf{x}_i$, $i = 1, \dots, N$.

In summary, we show below in Algorithm 1 the procedure of applying Mahalanobis distance detector for determining MCs in the data samples. Note that in step 4 T_D denotes the threshold on the distance $D(\mathbf{x}_i)$. By varying T_D , we can control the operating point in MC detection.

Algorithm 1

Mahalanobis distance detection algorithm

-
- 1: Obtain a set of NMC candidates $\mathcal{S} = \{\mathbf{x}_i, i = 1, \dots, N\}$.
 - 2: Estimate parameters μ and Σ according to (3).
 - 3: Compute the Mahalanobis distance $D(\mathbf{x}_i)$ for each sample $\mathbf{x}_i$.
 - 4: Determine $\mathbf{x}_i$ to be an outlier (i.e., MC) if $D(\mathbf{x}_i) > T_D$.
-

2.3. SNG outlier detection method

As in parametric detection methods, non-parametric methods for outlier detection also compute a score to indicate the presence of an outlier (Gao and Tan, 2006). Such an outlier score can be unbounded in value and may not necessarily be consistent from one dataset to another (Kriegel et al., 2009). In this work, we consider a stochastic outlier detection method introduced in Janssens (2013), which determines its output score in the form of a probability for a data sample to be an outlier. This probability score is desirable for our problem at hand, because it can be interpreted in a consistent way from one case to another.

The basic idea of the outlier detection method in Janssens (2013) is to represent the set of data samples $\{\mathbf{x}_i, i = 1, \dots, N\}$ by a so-called stochastic neighbor graph (SNG). This graph is formed by a set of nodes, to which the data samples $\mathbf{x}_i$ are mapped, and a set of directional edges, which reflect the affinity (or similarity) among the data samples. As a result, when a given sample is more similar to other samples, its corresponding node will have more edges directed to other nodes in the graph. In the end, the outliers are identified as those nodes having no edges pointing to them.

Specifically, let $G = (V, \mathcal{E}_G)$ denote the SNG formed by the set of data samples, where V is the set of nodes, i.e., $V = \{\mathbf{x}_i, i = 1, \dots, N\}$, and $\mathcal{E}_G$ is the set of directed edges connecting the nodes. The directed edges in $\mathcal{E}_G$ are generated according to a stochastic binding procedure as follows. For a given node $\mathbf{x}_i \in V$, it binds to each of the other nodes in V according to a prescribed probability distribution. When node $\mathbf{x}_i$ binds to node $\mathbf{x}_j$, a directed edge from $\mathbf{x}_i$ to $\mathbf{x}_j$ denoted by $i \rightarrow j$, is added to the graph. That is, $i \rightarrow j \in \mathcal{E}_G$.

By definition, the outliers in the data samples correspond to those nodes in G which no other nodes bind to. Let O denote the set of resulting outliers. Then the probability for node $\mathbf{x}_i$ to be an outlier can be written as:

$$P(\mathbf{x}_i \in O) = P(j \rightarrow i \notin \mathcal{E}_G, \forall \mathbf{x}_j \in V, j \neq i) = \prod_{j \neq i} [1 - P(j \rightarrow i \in \mathcal{E}_G)] \quad (7)$$

where $P(j \rightarrow i \in \mathcal{E}_G)$ denotes the probability of node $\mathbf{x}_j$ binding to node $\mathbf{x}_i$. The outlier probability in (7) is then used as the outlier score for sample $\mathbf{x}_i$ in the detection algorithm.

The binding probability $P(j \rightarrow i \in \mathcal{E}_G)$ above is defined according to an affinity measure between the two nodes (Janssens, 2013), which is given by:

$$a_{i,j} = \begin{cases} e^{-\|\mathbf{x}_i - \mathbf{x}_j\|/2\sigma_i^2} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (8)$$

where σ_i^2 is a variance parameter associated with $\mathbf{x}_i$. The binding probability is given by

$$P(j \rightarrow i \in \varepsilon_G) = \frac{a_{i,j}}{\sum_{m=1}^N a_{i,m}} \quad (9)$$

From the above definition it is seen that a node $\mathbf{x}_j$ is more likely to bind to $\mathbf{x}_i$ when the difference between their associated features is small. In (8), $\mathbf{x}_j$ is noted to have no affinity with itself. This is out of consideration that the outlier probability of a sample is influenced only by its similarity to samples other than itself.

In (8), the variance parameter σ_i^2 is solved independently for each node to control the sensitivity of the affinity measure to the difference between data samples. It is determined from the effective number of neighbors of $\mathbf{x}_i$, which is also known as a perplexity parameter h (a function of the entropy of the binding probability distribution (9)) (Goldberger et al., 2004). In our implementation, h was set to be $0.9 \times N$ as in Janssens (2013), and the parameter σ_i^2 was set such that node $\mathbf{x}_i$ distributes 90% of its total affinity measures into its h nearest neighbors in the graph. With the affinity measures distributed among such a large number of neighbors, the resulting outlier probability can be more robust to the presence of noisy data samples (Janssens, 2013).

In summary, we show in Algorithm 2 the procedure of applying the SNG method for determining the MCs in the data samples. Note that in step 4 T_O denotes the decision threshold for determining a sample to be an outlier or not. By varying T_O , we can control the operating point in MC detection.

Algorithm 2

SNG MC detection algorithm

-
- 1: Obtain a set of N MC candidates $\mathcal{S} = (\mathbf{x}_i, i = 1, \dots, N)$.
 - 2: For each sample $\mathbf{x}_i$, compute the binding probability distribution in (9) and determine the variance parameter σ_i^2 .
 - 3: Compute the outlier probability $P(\mathbf{x}_i \in \mathcal{O})$ in (7).
 - 4: Determine $\mathbf{x}_i$ to be an outlier (i.e., MC) if $P(\mathbf{x}_i \in \mathcal{O}) > T_O$.
-

2.4. Implementation of the outlier detection methods

2.4.1. Detection of candidate samples—To exploit the local noise characteristics in a given MC lesion $\mathcal{A}$, we considered the detector output in a surrounding region $\mathcal{R}$ of the lesion, as illustrated in Fig. 2. Given that the spatial extent of an MC cluster is typically less than 1 cm in diameter, we set $\mathcal{R}$ (which contains the lesion $\mathcal{A}$) to have a constant area 10 cm² for each lesion. Such a choice of region $\mathcal{R}$ was for it to be large enough to capture the local noise characteristics in the lesion area.

To detect the MCs in $\mathcal{A}$, the MC detector output in $\mathcal{R}$ is first compared against a threshold T to obtain a set of candidates $\{\mathbf{x}_i, i = 1, \dots, N\}$. These detected candidates also include detections in $\mathcal{A}$, among which both FPs and TPs can be present. We then apply our outlier detection approach to identify the TPs from them.

For a given lesion, it is desired to have all the MCs detected, but this requires the threshold T to be set very low, which will lead to a large number of candidates N . However, if the value of N is too large, the numerical complexity in the outlier detection procedure will unduly increase. Given that the number of MCs in a cluster is typically far less than 100, we selected the top $N = 600$ candidates for each lesion. These candidates were selected according to detector output strength among the detected objects (measured by the third largest pixel value within each object). These 600 candidates were then used as input to the outlier detection algorithm.

To reduce the impact on the outlier detection when the number of MCs is large in a lesion, we employed the following strategy for the SNG outlier detector: the set of $N = 600$ candidates is first divided into two subsets, denoted by A and B, respectively. Subset A consists of the top 100 candidates with the highest detector output, while subset B consists of the rest (potentially all FP candidates). Afterward, each of the candidates in A is tested in turn against the samples in B for outlier scoring. For the Mahalanobis detector, the set of $N = 600$ candidates is directly used, because the impact of outlier samples is alleviated through the use of the kernel function in (2). For MC detection, only the candidates in subset A were considered. This can save the computation time by ignoring those weaker candidates in B.

2.4.2. Vector description of candidates—As observed earlier in Fig. 1, the detector output is notably higher in a small neighborhood around an MC location, and the MC signal tends to be stronger than at the potential FP locations. To quantify the detector output, we used the detector output values in a window centered around each candidate in $\{\mathbf{x}_i, i = 1, \dots, N\}$. Considering that the average size of an MC is about 0.3–0.5 mm in diameter (Cheng et al., 2003), we set this window size to be 3×3 pixels at resolution 100 $\mu\text{m}/\text{pixel}$. Note that this image window is extracted in the detector output, wherein the extent of an MC is suppressed by the detector. This yields $k = 9$ feature components for the data samples $\mathbf{x}_i$. As an example, we show in Fig. 3 the image windows of the candidates obtained from the detector output earlier in Fig. 1; for better visualization, only a portion is shown for the ROIs.

For the SNG outlier detector, the feature vector $\mathbf{x}_i$ was further normalized among the data samples $\{\mathbf{x}_i, i = 1, \dots, N\}$ such that each feature component had zero mean and unit

standard-deviation as in Janssens (2013). For the Mahalanobis outlier detector, however, this normalization is not necessary because of the inherent normalization nature by the covariance matrix in (1).

3. Performance evaluation

3.1. Detection accuracy and case-to-case variability

To demonstrate the proposed locally adaptive decision approach in MC detection, we assessed its detection accuracy on a set of lesions containing clustered MCs. For each lesion, we examined both the number of true MCs detected and the number of FPs. To summarize the overall performance, we used free response operating characteristic (FROC) curve, which is a plot of the true-positive fraction (TPF) vs the number of FPs over the continuum of the operating threshold. The TPF (aka sensitivity) is defined as the number of detected true MCs divided by the total number of true MCs in a lesion. A higher FROC curve corresponds to better detection performance.

In addition, we also evaluated the case-to-case variability in the detection results. For this purpose, we examined the variations in the level of FPs among the different lesions when the same MC detector was operated at given sensitivity level. This variability was quantified by using the standard-deviation of the number of FPs detected among different lesions.

To quantify the FROC results, we calculated the partial area under the FROC curve (pAUC) (Walter, 2005). A bootstrapping procedure with 20,000 samples was applied in order to reduce the variations associated with case distribution (Samuelson and Petrick, 2006), based on which statistical comparisons are made on the detection performance by different methods.

3.2. Mammogram dataset

We made use of a set of full-field digital mammography (FFDM) images collected by the Department of Radiology of the University of Chicago (U of C) under IRB approval. They were acquired using a Senographe 2000D FFDM system (General Electric Medical Systems; Milwaukee, WI) with a pixel size of $100\text{ }\mu\text{m/pixel}$. The set consisted of 188 images from 95 cases (43 malignant, 52 benign), all containing clustered MCs. They were collected consecutively over time and were all sent for biopsy due to the subtlety of their MC lesions (patient age: 59.6 ± 10.5). Most of the cases had both craniocaudal (CC) and mediolateral (MLO) oblique views as well. The distribution of these mammograms among different BIRADS categories is given in Table 1.

The individual MCs in the mammogram images were marked by a researcher with more than 15 years of experience in mammography research and with special training on interpreting mammograms and reviewed by a researcher with more than 5 years of experience in mammography research. There were a total of 8,979 MCs marked, which were used as ground truth for evaluation.

For the purpose of evaluating the detection accuracy, the lesion regions of clustered MCs in these mammograms were marked out by a bounding circle with a diameter of 1 cm, 2 cm, or

3 cm (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. The detection results (both FPs and TPs) were then assessed for each of these lesion regions.

3.3. MC detectors for demonstration

To demonstrate the proposed approach, we considered two well-cited MC detectors in the literature. The first is the SVM detector developed in El-Naqa et al. (2002), which is based on machine learning; the second is the DoG detector (Dengler et al., 1993), which is of low computational complexity and based on image enhancement for MC detection. These two detectors differ in their detection performance, as to be seen in the results, thus providing a test bed for the proposed approach with different detection accuracy levels.

To determine the detected MCs in a given image, the detector output was first compared against an operating threshold T at each pixel. Afterward, the pixels above the threshold were grouped to form objects with 8-neighborhood connectedness. A detected object with 2 pixels or less in size was eliminated from further consideration in order to avoid spurious detections. In the end, a detected object was treated as a TP when it was less than 0.3 mm away from a marked MC or at least 40% of its area overlapped with that of an MC; otherwise it was counted as an FP.

For reducing the effect of linear structures in the breast, which are known to cause FPs in MC detection (Ema et al., 1995; Zwiggelaar et al., 2004), we applied a bithresholding approach as in (Wang et al., 2013). In this approach, the linear structures were first separated from the image region under consideration by using a line detector. The MC detection threshold T was then adjusted for the linear structures according to the mean and standard deviation in the detector output such that it was equivalent to the threshold level in the rest of the image region (Wang et al., 2013).

4. Results

Below we first present results on the detection accuracy achieved by the proposed outlier decision approach. Afterward, we present results to demonstrate the case-to-case variability in the detection results. For comparison, we also provide results obtained by the traditional thresholding approach in MC detection, where the detector output is compared against a common threshold for all cases ('Uniform method'). The uniform thresholding approach is most commonly used among the MC detectors in the literature (e.g. Zhang et al., 2014). Here it is used as a baseline in detection decision, and our purpose is to demonstrate whether the proposed adaptive decision approach can lead to any improvement in performance for a given MC detector. For convenience, we denote the Mahalanobis distance outlier detection method by MD, and the stochastic neighbor graph method by SNG.

4.1. Detection accuracy by outlier decision approach

4.1.1. SVM detector—To summarize the detection performance, we show in Fig. 4 the FROC curves obtained by the proposed outlier decision approach (MD and SNG). For comparison, the FROC curve obtained with the traditional thresholding approach is also

given (Uniform). In Fig. 4, the y-axis represents the TPF (i.e., sensitivity) in detection, whereas the x-axis represents the number of FPs detected per unit area of the lesion (equivalent to a circle with 1 cm diameter).

It can be seen that the FROC curves of both outlier methods MD and SNG are notably higher than that of traditional thresholding. At a given FP rate, the outlier approach yields a higher sensitivity in detection. For example, with FPs at 2 per lesion area, the sensitivity values of MD and SNG are 0.70 and 0.72, respectively, compared to 0.65 for Uniform (p -values $< 10^{-4}$ for both MD and SNG). On the other hand, at a given sensitivity level, the outlier approach yields a lower FP rate in detection. For example, with sensitivity at 70%, MD and SNG achieved 1.94 and 1.63 FPs per unit area, respectively, compared to 3.19 FPs for Uniform (p -value $< 10^{-4}$ for both MD and SNG).

Furthermore, a statistical comparison of the pAUC values of the FROC curves between MD and Uniform yields 4.21 vs. 3.95 (p -value $< 10^{-4}$); similarly, the pAUC value for SNG is 4.25 (p -value $< 10^{-4}$ compared to Uniform). The pAUC value of SNG is also noted to be slightly higher than that of MD (p -value $< 10^{-4}$). In these results, the pAUC was calculated over the FP rate interval from 0 to 6 FPs per lesion area.

As an example, we show in Fig. 5 the detection results by the different methods for the two example ROIs earlier in Fig. 1. The sensitivity was set at 70% for all the decision methods. In the first lesion, six out of the seven marked MCs are detected by all the three decision methods, but there are four FPs in Uniform and only one in both MD and SNG. In the second lesion, all of the marked MCs are detected by the three methods, but the number of FPs is nine in Uniform, none in MD, and one in SNG.

4.1.2. DoG detector—In Fig. 6 we show the FROC curves obtained by the outlier approach (MD and SNG) when applied to the output of the DoG detector. For comparison, the FROC curve is also given for the conventional thresholding approach (Uniform). As in the results of the SVM detector above, the FROC curves of the outlier methods are higher than that of uniform thresholding. Specifically, at 2 FPs per lesion area, the sensitivity values of MD and SNG are both 0.69, compared to 0.61 for Uniform (p -value $< 10^{-4}$ for both MD and SNG). At sensitivity level of 70%, the FP rates are 2.42, 2.43 and 3.97 FPs per unit lesion area, respectively, for MD, SNG and Uniform (p -value $< 10^{-4}$ for both MD and SNG vs. Uniform).

In addition, a statistical comparison of the pAUC values of the FROC curves yields 4.09 and 4.07 for MD and SNG, compared to 3.77 for Uniform (p -value $< 10^{-4}$ for both MD and SNG). However, there was no significant difference between the two outlier methods (p -value = 0.223); a 95% confidence interval for the difference in pAUCs is $[-0.007, 0.051]$.

In Fig. 7 we show the detection results for the example ROIs shown earlier in Fig. 1. Given the higher FP rate in the DoG detector, the sensitivity was set at 65%. In the first lesion, there are six MCs (out of seven) detected by all the methods, but three FPs in Uniform, one FP in SNG, and none in MD. In the second lesion, all of the marked MCs are detected by the

three methods, but there are five FPs in Uniform, compared to only one FP in both SNG and MD.

4.2. Case-to-case variability in MC detection

4.2.1. SVM detector—The FROC results above demonstrate that the outlier approach can yield higher detection accuracy on average than the traditional uniform decision approach. We also examined the variability of the detection accuracy among the detection results of different cases. In Fig. 8 we show a plot of the level of FPs detected by the different methods at sensitivity levels of 60%, 65% and 70%. At each sensitivity level, both the mean and standard-deviation of the number of FPs (per unit area) among different lesions in the dataset are shown; the error bars correspond to the standard-deviation values. As can be seen, the variability is notably reduced in the outlier approach. For example, at sensitivity 60%, the standard-deviation values are 0.86 and 0.84 for MD and SNG, respectively, compared to 1.54 for Uniform. A lower standard-deviation value implies a more uniform level of FPs detected among different lesions.

4.2.2. DoG detector—Similarly, we show in Fig. 9 a plot of the FP levels obtained by the different methods when applied to the output of the DoG detector. In the plot, the FP levels are shown for the following three sensitivity levels: 55%, 60% and 65%. Note that the FP level is much higher in DoG than in SVM. Thus, we have reduced the sensitivity range in order to maintain a similar FP level as in SVM above. For this reason, we choose sensitivity levels in this operating range. As can be seen, the standard-deviation values of FPs are lower in both MD and SNG than in Uniform. In particular, at sensitivity 55%, the standard-deviation values are 0.75 and 0.85 for MD and SNG, respectively, compared to 1.95 for Uniform.

5. Discussions

5.1. Comparison of the outlier methods

The evaluation results above demonstrate that the proposed outlier decision approach could yield improved accuracy in MC detection over the traditional approach of uniform thresholding. Interestingly, the two outlier decision methods MD and SNG yielded comparable results in the DoG detector output. The FROC curves in Fig. 6 showed no statistical difference between the two methods. However, for the SVM detector, SNG yielded a slightly higher FROC curve than MD in Fig. 4. For example, at 1.5 FPs per lesion, the sensitivity of SNG is 0.69, compared to 0.68 for MD. This difference is likely due to that SNG is a non-parametric approach and as a result it could better adapt to the noise characteristics in the SVM detector. Nevertheless, the absolute difference between the two FROC curves is rather small.

In terms of computational complexity, in the MD decision method, we need to estimate a total of 54 parameters (9 for μ and 45 for Σ) for dimension $k = 9$ from the data samples. The Mahalanobis distance also needs to be computed for each of the 100 samples in subset A. On the other hand, the SNG decision method requires not only computing the pairwise distances among the 500 samples in subset B, but also determining the variance parameter σ_i^2

in (8) for each sample. This is more expensive computationally than the MD method. Of course, the SNG method can be more flexible in that it does not assume a parametric distribution on the MC detector output.

5.2. Use of human marked MCs as ground truth

In this study we quantified the detection accuracy by the different methods based on human marked MCs. In the literature, human marked MCs are routinely used as the ground truth in development of MC detection algorithms (Cheng et al., 1998; Freer and Ullissey, 2001; Balakumaran et al., 2010; Yao et al., 2012). However, human marked MCs are inevitably subject to errors associated with inter- and intra-observer variations. As a result, the obtained performance could be affected by the errors associated with the marked MCs. Our main results indicate that at a given sensitivity level the proposed decision strategy could achieve a lower FP rate on average and reduce the variability of FP rate among different cases. We do not expect the significance of these results to change as long as the level of errors in the marked MCs is reasonably low, because the evaluation is averaged over such a large number of MCs. Indeed, in our previous studies (Sainz de Cea et al., 2017), the marked MCs in the same dataset yielded the best accuracy when used for classifying their associated lesions as malignant or benign when compared with MCs detected by a computerized detector; moreover, those cases with detected MCs closer to the marked MCs also yielded higher classification accuracy than those with larger differences from the marked MCs.

Nevertheless, to validate these main results, we also evaluated the detection results based on the marked MCs from an additional reader. Given the larger number of MCs, we randomly selected 30 cases (60 mammogram images) from the dataset, and had these cases marked by a second expert reader. We then conducted an analysis of the detection results based on this second set of marked MCs. The results were consistent with that the proposed methods yielded a lower FP rate and reduced case-to-case variability. Specifically, when the SVM detector was used, the number of FPs per unit area at sensitivity level 75% was 1.43 ± 1.34 for SNG, compared to 2.81 ± 2.30 for the uniform method (p -value $< 10^{-4}$); the number of FPs per unit area was 1.55 ± 1.49 for MD (p -value $< 10^{-4}$ compared to the uniform method). Similarly, when the DoG detector was used, the number of FPs per unit area at sensitivity level 70% was 1.25 ± 1.13 for SNG, compared to 1.86 ± 1.94 for the uniform method (p -value $< 10^{-4}$); the number of FPs per unit area was 1.00 ± 1.15 for MD (p -value = 0.015 compared to the uniform method).

5.3. Limitations and future work

For quantifying the noise characteristics in a lesion region $\mathcal{A}$, we considered the detector output in a fixed image region $\mathcal{R}$ surrounding the lesion. However, the image characteristics may not be stationary in $\mathcal{R}$ depending on the local breast tissue. This may adversely affect the outlier decision outcome, because the detected noise samples in $\mathcal{R}$ may now have different statistical properties as in the lesion region $\mathcal{A}$. In future work, it would be worthy to further refine the region $\mathcal{R}$ (e.g., according to local tissue density) such that its characteristics are more similar to that of $\mathcal{A}$.

In addition, in this work we considered two existing outlier detection methods for our task. It would be interesting to further investigate using some more advanced methods in outlier detection, such as the one-class SVM classifier which has been successfully applied in document classification (Manevitz and Yousef, 2001), image retrieval (Chen et al., 2001), etc.

6. Conclusion

We investigated an adaptive decision approach in detection of individual MCs in mammogram lesions, and demonstrated its performance with two existing MC detectors, namely the SVM and DoG detectors. Instead of developing a new MC detector, the proposed approach is on how to better decide whether a detected object is an MC or not in the detector output. In our formulation, we treat the MCs as statistical outliers compared among the many noisy detections in the detector output in the local lesion region. To identify the MCs, we considered two statistical outlier detection methods, one parametric and one non-parametric. For the parametric method, we adopted the Mahalanobis distance detector which is based on the assumption of a multi-dimensional Gaussian distribution on the inlier data samples (i.e. potential FPs). For a given lesion image, the parameters of this distribution were estimated from the detected objects using kernel weighting in order to mitigate the effect of the MCs present (whose locations are unknown) within the lesion. The MCs were then identified as those samples having the largest Mahalanobis distance values from the underlying distribution. In the non-parametric method, no explicit distribution was assumed for either the FP or the TP class. Instead, it used a stochastic neighbor graph to model the similarity or dissimilarity among the individual data samples. The outliers were determined to be those samples which the rest of the data samples less likely bind to. The binding probabilities among the samples were derived from their mutual distances in the feature space. For a sample under consideration, the outlier probability was computed from the stochastic neighbor graph. We demonstrated the proposed outlier decision approach on a dataset of 188 FFDM images from 95 cases. The results show that the proposed approach could not only reduce the number of FPs at a given sensitivity level, but also reduce the variability among different lesions.

Acknowledgments

This work was supported partially by NIH/NIBIB under grant R01EB009905.

References

1. American Cancer Society. Cancer Facts and Figures. Atlanta, GA: 2017. (n.d.).
2. Apostolopoulos, G., Koutras, A., Christoyianni, I., Dermatas, E. Hellenic Conference on Artificial Intelligence. Springer; 2014. Computer aided classification of mammographic tissue using shapelets and support vector machines; p. 510-520.
3. Balakumaran T, Vennila I, Shankar CG. Detection of microcalcification in mammograms using wavelet transform and fuzzy shell clustering. arXiv preprint arXiv:1002.2182. 2010
4. Bowman, AW., Azzalini, A. Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations. Vol. 18. OUP Oxford; 1997.
5. Branch JW, Giannella C, Szymanski B, Wolff R, Kargupta H. Innetwork outlier detection in wireless sensor networks. Knowledge and Information Systems. 2013; 34(1):23–54.

6. Bria A, Karssemeijer N, Tortorella F. Learning from unbalanced data: a cascade-based approach for detecting clustered microcalcifications. *Medical Image Analysis*. 2014; 18(2):241–252. [PubMed: 24292553]
7. Cao, L., Yang, D., Wang, Q., Yu, Y., Wang, J., Rundensteiner, EA. International Conference on Data Engineering (ICDE). IEEE; 2014. Scalable distance-based outlier detection over high-volume data streams; p. 76–87.
8. Chen, Y., Zhou, XS., Huang, TS. International Conference on Image Processing (ICIP). Vol. 1. IEEE; 2001. One-class SVM for learning in image retrieval; p. 34–37.
9. Cheng HD, Cai X, Chen X, Hu L, Lou X. Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recognition*. 2003; 36(12):2967–2991.
10. Cheng HD, Lui YM, Freimanis RI. A novel approach to microcalcification detection using fuzzy logic technique. *IEEE Transactions on Medical Imaging*. 1998; 17(3):442–450. [PubMed: 9735907]
11. Dengler J, Behrens S, Desaga JF. Segmentation of microcalcifications in mammograms. *IEEE Transactions on Medical Imaging*. 1993; 12(4):634–642. [PubMed: 18218457]
12. Desclée B, Bogaert P, Defourny P. Forest change detection by statistical object-based method. *Remote Sensing of Environment*. 2006; 102(1):1–11.
13. Eguchi S, Copas J. A class of local likelihood methods and nearparametric asymptotics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 1998; 60(4):709–724.
14. El-Naga I, Yang Y, Wernick MN, Galatsanos NP, Nishikawa RM. A support vector machine approach for detection of microcalcifications. *IEEE Transactions on Medical Imaging*. 2002; 21(12):1552–1563. [PubMed: 12588039]
15. Elter M, Horsch A. CADx of mammographic masses and clustered microcalcifications: a review. *Medical Physics*. 2009; 36(6):2052–2068. [PubMed: 19610294]
16. Ema T, Doi K, Nishikawa RM, Jiang Y, Papaioannou J. Image feature analysis and computer-aided diagnosis in mammography: Reduction of false-positive clustered microcalcifications using local edge-gradient analysis. *Medical Physics*. 1995; 22(2):161–169. [PubMed: 7565347]
17. Filzmoser P. Identification of multivariate outliers: a performance study. *Austrian Journal of Statistics*. 2016; 34(2):127–138.
18. Filzmoser P, Garrett RG, Reimann C. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences*. 2005; 31(5):579–587.
19. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center 1. *Radiology*. 2001; 220(3):781–786. [PubMed: 11526282]
20. Gao, J., Tan, PN. Sixth International Conference on Data Mining, 2006. ICDM'06. IEEE; 2006. Converting output scores from outlier detection algorithms into probability estimates; p. 212–221.
21. Glenn N, Zhao Y. Weighted empirical likelihood estimates and their robustness properties. *Computational Statistics & Data Analysis*. 2007; 51(10):5130–5141.
22. Goldberger J, Hinton GE, Roweis ST, Salakhutdinov R. Neighbourhood components analysis. *Advances in Neural Information Processing Systems*. 2004:513–520.
23. Gunawan, D. Communications, Computers and signal Processing, 2001 PACRIM 2001 IEEE Pacific Rim Conference on. Vol. 2. IEEE; 2001. Microcalcification detection using wavelet transform; p. 694–697.
24. Gur D, Sumkin JH, Rockette HE, Ganott M, Hakim C, Hardesty L, Poller WR, Shah R, Wallace L. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *Journal of the National Cancer Institute*. 2004; 96(3):185–190. [PubMed: 14759985]
25. Gurcan MN, Yardimci Y, Cetin AE, Ansari R. Detection of microcalcifications in mammograms using higher order statistics. *IEEE Signal Processing Letters*. 1997; 4(8):213–216.
26. Hyndman, RL., Zhang, X., King, ML., et al. Econometric Society 2004 Australasian Meetings. Econometric Society; 2004. Bandwidth selection for multivariate kernel density estimation using MCMC.

27. Jalalian A, Mashohor SB, Mahmud HR, Saripan MIB, Ramli ARB, Karasfi B. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clinical Imaging*. 2013; 37(3):420–426. [PubMed: 23153689]
28. Janssens, J. PhD thesis. Tilburg University; 2013. Outlier selection and one-class classification.
29. Jiang MF, Tseng SS, Su CM. Two-phase clustering process for outliers detection. *Pattern Recognition Letters*. 2001; 22(6):691–700.
30. Jing H, Yang Y, Nishikawa RM. Detection of clustered microcalcifications using spatial point process modeling. *Physics in Medicine and Biology*. 2011; 56(1):1–17. [PubMed: 21119233]
31. Karahaliou A, Skiadopoulos S, Boniatis I, Sakellaropoulos P, Likaki E, Panayiotakis G, Costaridou L. Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis. *The British Journal of Radiology*. 2014
32. Kriegel, HP., Kröger, P., Schubert, E., Zimek, A. Proceedings of the 18th ACM Conference on Information and Knowledge Management. ACM; 2009. LoOP: local outlier probabilities; p. 1649-1652.
33. Lancaster, HO., Seneta, E. Chi-Square Distribution. Wiley Online Library; 2005.
34. Linguraru MG, Marias K, English R, Brady M. A biologically inspired algorithm for microcalcification cluster detection. *Medical Image Analysis*. 2006; 10(6):850–862. [PubMed: 16945569]
35. Lukashevich, H., Nowak, S., Dunker, P. International Conference on Multimedia and Expo (ICME). IEEE; 2009. Using one-class SVM outliers detection for verification of collaboratively tagged image training sets; p. 682-685.
36. Mahalanobis PC. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*. 1936; 2:49–55.
37. Manevitz LM, Yousef M. One-class SVMs for document classification. *Journal of Machine Learning Research*. 2001 Dec.2:139–154.
38. Marsland, SR. On-line Novelty Detection Through Self-organisation, with Application to Inspection Robotics. University of Manchester; 2001.
39. McLoughlin KJ, Bones PJ, Karssemeijer N. Noise equalization for detection of microcalcification clusters in direct digital mammogram images. *IEEE Transactions on Medical Imaging*. 2004; 23(3):313–320. [PubMed: 15027524]
40. Miranda GHB, Felipe JC. Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization. *Computers in Biology and Medicine*. 2015; 64:334–346. [PubMed: 25453323]
41. Mordang JJ, Gubern-Mérida A, den Heeten G, Karssemeijer N. Reducing false positives of microcalcification detection systems by removal of breast arterial calcifications. *Medical Physics*. 2016; 43(4):1676–1687. [PubMed: 27036566]
42. Naseem M, Murray J, Hilton JF, Karamchandani J, Muradali D, Faragalla H, Polenz C, Han D, Bell DC, Brezden-Masley C. Mammographic microcalcifications and breast cancer tumorigenesis: a radiologic-pathologic analysis. *BMC Cancer*. 2015; 15(1):307. [PubMed: 25896922]
43. Oliver A, Torrent A, Lladó X, Tortajada M, Tortajada L, Sentís M, Freixenet J, Zwiggelaar R. Automatic microcalcification and cluster detection for digital and digitised mammograms. *Knowledge-Based Systems*. 2012; 28:68–75.
44. Pereira DC, Ramos RP, Do Nascimento MZ. Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm. *Computer Methods and Programs in Biomedicine*. 2014; 114(1):88–101. [PubMed: 24513228]
45. Prastawa M, Bullitt E, Ho S, Gerig G. A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis*. 2004; 8(3):275–283. [PubMed: 15450222]
46. Radovanovi M, Nanopoulos A, Ivanovi M. Reverse nearest neighbors in unsupervised distance-based outlier detection. *IEEE Transactions on Knowledge and Data Engineering*. 2015; 27(5): 1369–1382.
47. Ramaswamy, S., Rastogi, R., Shim, K. ACM SIGMOD Record. Vol. 29. ACM; 2000. Efficient algorithms for mining outliers from large data sets; p. 427-438.
48. Rangayyan RM, Ayres FJ, Desautels JL. A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. *Journal of the Franklin Institute*. 2007; 344(3):312–348.

49. Rousseeuw, PJ., Leroy, AM. Robust regression and outlier detection. Vol. 589. John Wiley & Sons; 2005.
50. Sainz de Cea MV, Nishikawa RM, Yang Y. Estimating the accuracy level among individual detections in clustered microcalcifications. *IEEE Transactions on Medical Imaging*. 2017; 36(5): 1162–1171. [PubMed: 28103550]
51. Sainz de Cea, MV., Yang, Y. *IEEE International Conference on Image Processing (ICIP)*. IEEE; 2015. Improving uniformity in detection performance of clustered microcalcifications in mammograms; p. 842-846.2015
52. Sainz de Cea, MV., Yang, Y. *International Conference on Image Processing (ICIP)*. IEEE; 2016. Case-based decision strategy using outlier probability in detection of microcalcifications in mammographic lesions.
53. Samuelson, FW., Petrick, N. 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006. IEEE; 2006. Comparing image detection algorithms using resampling; p. 1312-1315.
54. Schmidt P, Gaser C, Arsic M, Buck D, Förchler A, Berthele A, Hoshi M, Ilg R, Schmid VJ, Zimmer C, et al. An automated tool for detection of flair-hyperintense white-matter lesions in multiple sclerosis. *Neuroimage*. 2012; 59(4):3774–3783. [PubMed: 22119648]
55. Scholz F. Maximum likelihood estimation. *Encyclopedia of Statistical Sciences*. 1985
56. Schubert, E., Zimek, A., Kriegel, HP. *Proceedings of the 2014 SIAM International Conference on Data Mining*. SIAM; 2014. Generalized outlier detection with flexible kernel density estimates; p. 542-550.
57. Scimeca M, Giannini E, Antonacci C, Pistolese CA, Spagnoli LG, Bonanno E. Microcalcifications in breast cancer: an active phenomenon mediated by epithelial cells with mesenchymal characteristics. *BMC Cancer*. 2014; 14(1):286–295. [PubMed: 24758513]
58. Scott DW, Sain SR. Multidimensional density estimation. *Handbook of Statistics*. 2005; 24:229–261.
59. Tang J, Rangayyan RM, Xu J, El Naqa I, Yang Y. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Transactions on Information Technology in Biomedicine*. 2009; 13(2):236–251. [PubMed: 19171527]
60. Terrell GR, Scott DW. Variable kernel density estimation. *The Annals of Statistics*. 1992:1236–1265.
61. Thangavel, K., Mohideen, AK. *Trendz in Information Sciences & Computing (TISC)*, 2010. IEEE; 2010. Semi-supervised k-means clustering for outlier detection in mammogram classification; p. 68-72.
62. Tibshirani R, Hastie T. Local likelihood estimation. *Journal of the American Statistical Association*. 1987; 82(398):559–567.
63. Todeschini R, Ballabio D, Consonni V, Sahigara F, Filzmoser P. Locally centred Mahalanobis distance: a new distance measure with salient features towards outlier detection. *Analytica Chimica Acta*. 2013; 787:1–9. [PubMed: 23830416]
64. Van Leemput K, Maes F, Vandermeulen D, Colchester A, Suetens P. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*. 2001; 20(8):677–688. [PubMed: 11513020]
65. Walter S. The partial area under the summary ROC curve. *Statistics in Medicine*. 2005; 24(13): 2025–2040. [PubMed: 15900606]
66. Wang J, Nishikawa RM, Yang Y. Improving the accuracy in detection of clustered microcalcifications with a context-sensitive classification model. *Medical Physics*. 2016; 43(1): 159–170. [PubMed: 26745908]
67. Wang, J., Yang, Y., Nishikawa, RM. *International Conference on Image Processing (ICIP)*. IEEE; 2013. Reduction of false positive detection in clustered microcalcifications.
68. Yao, C., Yang, Y., Chen, H., Jing, T., Hao, X., Bi, H. *IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. IEEE; 2012. Adaptive kernel learning for detection of clustered microcalcifications in mammograms; p. 5-8.
69. Yu SN, Li KY, Huang YK. Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model. *Computerized Medical Imaging and Graphics*. 2006; 30(3): 163–173. [PubMed: 16723208]

70. Zhang E, Wang F, Li Y, Bai X. Automatic detection of microcalcifications using mathematical morphology and a support vector machine. *Bio-medical Materials and Engineering*. 2014; 24(1): 53–59.
71. Zhang Y, Du B, Zhang L, Wang S. A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*. 2016; 54(3):1376–1389.
72. Zhou X, Yang C, Yu W. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2013; 35(3):597–610. [PubMed: 22689075]
73. Zwiggelaar R, Astley SM, Boggis CR, Taylor CJ. Linear structures in mammographic images: detection and classification. *IEEE Transactions on Medical Imaging*. 2004; 23(9):1077–1086. [PubMed: 15377116]

Figure 1.

Example mammogram ROIs from different cases with MCs (a)/(c), and their MC detector output (b)/(d). The MCs are indicated by blue arrows in (a)/(c). The area is 163×163 pixels ($1.63 \text{ cm} \times 1.63 \text{ cm}$) in both ROIs. The contrast of the images was adjusted for better visualization of the MCs.

Figure 2.

Example of a lesion region $\mathcal{A}$ and its surrounding region $\mathcal{R}$. The MCs are indicated by green circles. The size of the ROI is $4.7 \text{ cm} \times 4.7 \text{ cm}$.

Figure 3.
Image windows of the candidates in the SVM detector output for the ROIs in Fig. 1.

Figure 4. FROC curves obtained by the following different decision methods for the SVM detector: proposed outlier methods MD and SNG, and traditional uniform thresholding (Uniform). The Y-axis represents the sensitivity level obtained at a corresponding FP rate.

Figure 5.

Examples of detection results by the three decision methods in the SVM detector for the ROIs in Fig. 1. The MCs are marked by green circles and the detected MCs by red x's. The left column is for ROI 1, and the right column is for ROI 2.

Figure 6.

FROC curves obtained by the following different decision methods for the DoG detector: proposed outlier methods MD and SNG, and traditional uniform thresholding (Uniform). The Y-axis represents the sensitivity level obtained at a corresponding FP rate.

Figure 7.

Examples of detection results by the three decision methods in the DoG detector for the ROIs in Fig. 1. The MCs are marked by green circles and the detected MCs by red x's. The left column is for ROI 1, and the right column is for ROI 2.

Figure 8.

The mean and standard deviation of the number of FPs per lesion area at different sensitivity levels for the SVM detector.

Figure 9.

The mean and standard deviation of the number of FPs per lesion area at different sensitivity levels for the DoG detector.

Table 1

Distribution of the mammograms in the dataset among different BIRADS categories

BIRADS category	1	2	3	4
No. of mammograms	35	79	60	14