A context-sensitive deep learning approach for microcalcification detection in mammograms

Juan Wang[a], Yongyi Yang[a,*]

[a] *Medical Imaging Research Center, Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616*

Abstract

A challenging issue in computerized detection of clustered microcalcifications (MCs) is the frequent occurrence of false positives (FPs) caused by local image patterns that resemble MCs. We develop a context-sensitive deep neural network (DNN), aimed to take into account both the local image features of an MC and its surrounding tissue background, for MC detection. The DNN classifier is trained to automatically extract the relevant image features of an MC as well as its image context. The proposed approach was evaluated on a set of 292 mammograms using free-response receiver operating characteristic (FROC) analysis on the accuracy both in detecting individual MCs and in detecting MC clusters. The results demonstrate that the proposed approach could achieve significantly higher FROC curves when compared to two MC-based detectors. It indicates that incorporating image context information in MC detection can be beneficial for reducing the FPs in detections.

Keywords: Computer-aided diagnosis (CAD), clustered microcalcifications (MCs), deep neural network (DNN), deep learning.

1. Introduction

Breast cancer is currently the most frequently diagnosed non-skin cancer in women. It is estimated that about 252,710 new breast cancer cases and 40,610

*Corresponding author
Email address: yy@ece.iit.edu (Yongyi Yang)
[1] This work was supported by NIH/NIBIB under grant R01EB009905.

breast cancer deaths will occur among women in the US in 2017 [1]. Mammography is an effective screening tool for diagnosis of breast cancer, which can detect about 80%-90% of breast cancer cases in women without symptoms [1]. One important early sign of breast cancer in mammograms is the appearance of clustered microcalcifications (MCs), which are found in 30%-50% of mammographically diagnosed cancer cases [2, 3]. MCs are tiny calcium deposits which appear as bright spots in mammograms (Figure 1). Clustered MCs are formed by a group of MCs closely distributed within a spatial region (typically within a 1 cm^2 area). While often seen, accurate detection and diagnosis of MCs in mammograms can be difficult, because of their subtlety in appearance, variation in shape and size, and inhomogeneity in surrounding tissue. Clustered MCs can occur both in benign cases and in malignant cases. Compared with malignant cases, benign MC clusters tend to have fewer and smaller MCs; on the other hand, clustered MCs in malignant cases tend to exhibit a wider variation in appearances within a cluster [4].

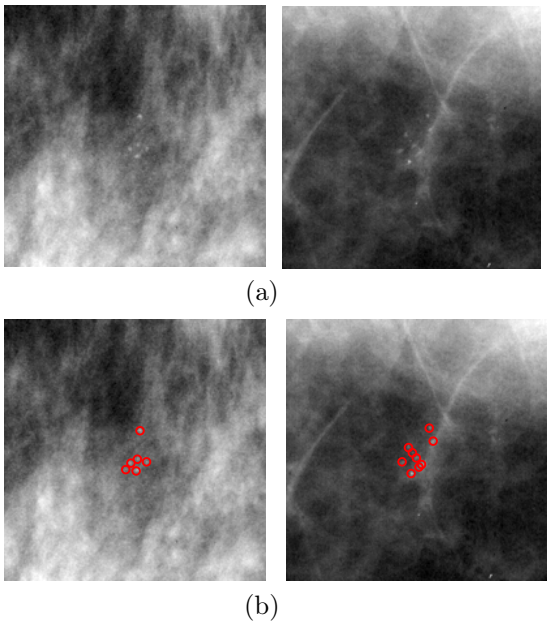


Figure 1: (a) Example ROIs with clustered MCs (200×200 pixels, 0.1 mm/pixel); (b) Locations of individual MCs marked by red circles.

In the literature, computerized methods have been investigated for accurate detection of clustered MCs [5, 6, 7, 8, 9, 10]. These methods are referred to as computer-aided detection (CADe) in screening mammography, the purpose of which is to serve as an alert to the radiologists by detecting the presence of suspicious regions in a mammogram for examination. In these methods, detection of clustered MCs is typically performed in two steps. In the first step, an MC detector is applied to locate the candidates of individual MCs in a mammogram; afterward, the detected MCs are grouped into clusters according to a set of clustering criteria. Based on the fact that individual MCs are small (typically 0.1-1 mm in diameter), most of the detectors (if not all) are designed to exploit the local image characteristics of an MC (e.g., [5, 6, 8, 10, 11]). For example, Salfity *et al* [5] used a difference-of-Gaussians (DoG) filter wherein the filter consisted of two kernels of limited width parameters. El-Naqa *et al* [6] adopted a local image window centered at a location under consideration. Oliver *et al* [8] extracted features by using a filter bank to obtain a description of the local morphology of an MC. Mordang *et al* [10] applied two CNNs for MC detection in multi-vendor mammography, in which one CNN was used to remove easy samples and another was used for classifying the survived samples; the input sample consisted of a local window at a detection location. Samala *et al* [11] studied the use of CNNs with different parameter settings to reduce false positives (FPs) in MC detection in digital breast tomosynthesis.

While successful in achieving high sensitivity, a challenging issue facing MC detection methods is the frequent occurrence of FPs. This is because the detector response is susceptible to local image patterns that resemble MCs. Indeed, there are several known factors that can contribute to the occurrence of FPs in mammograms, including MC-like noise patterns, linear structures (attributed to vessels, ducts, fibrous tissue, skinfolds, edges and other anatomical features), inhomogeneity in tissue background, imaging artifacts, etc [12]. There have been studies on how to suppress FPs in MC detection. These methods typically exploit not only the image characteristics of the MCs themselves but also their surroundings. For example, noise equalization techniques were developed

for reducing the noise variability in a mammogram [13, 14]; background removal methods were used to suppress the inhomogeneity in tissue background [15]; image features associated with linear structures were also incorporated for reducing FPs [16, 17, 12, 9].

Recently, we developed a different approach in [18] to detecting the presence of clustered MCs in mammograms. Instead of first detecting the MCs individually, we applied a convolutional neural network (CNN) to determine directly whether an image region contains an MC cluster or not. With this approach, the input to the CNN classifier was formed by a large image window (roughly 1 cm^2 in area) that contains not only the individual MCs but also their surrounding image context. The CNN classifier was then optimized through training to learn the relevant features for detecting MC clusters. This approach was demonstrated to be more robust to the FP patterns in mammograms when compared to several existing MC detectors [18]. This detector is intended only for identifying suspicious regions for subsequent examination. It does not specify the locations of individual MCs within the detected region.

Built on the success of the global detection approach in [18], in this study we investigate the feasibility of extending this approach by also incorporating local MC features such that it could improve the accuracy in detecting individual MCs. In computer-aided diagnosis (CADx), accurately detecting the individual MCs in a cluster is important, because the image features of the detected MCs are further analyzed for classification as being benign or malignant [19, 20, 21, 22, 23]. Studies have shown that the accuracy of detected individual MCs can impact on the CADx performance [24, 25, 26].

We develop a deep neural network (DNN) architecture, aimed to take into account not only the local image features of an MC but also its surrounding image context, for MC detection. Specifically, the detector network is formed by two subnetworks, one for extracting the local image features and one for learning the image features of its surrounding background. The extracted features by the two subnetworks are combined subsequently for classifying whether an MC is present or not at a detection location. Consequently, the detector response

is automatically adapted to the image background at an MC; the proposed detector is termed as context-sensitive DNN accordingly. In the experiments, we evaluated the proposed approach both for the task of correctly identifying individual MCs in a given image region and for the task of detecting the presence of MC clusters in mammograms. Both tasks are essential in developing CADe and CADx systems for clustered MCs.

We note that deep learning methods have been increasingly applied to a variety of problems in medical imaging in recent years. For example, in [27] Inception-v3 networks were used for detecting diabetic retinopathy in retinal fundus photographs. In [28] an ensemble of CNNs was applied to discriminate between benign and malignant breast tumors in breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). In [29] a parasitic metric learning net was used for breast mass classification in mammograms. In [30] a deep Siamese neural network was studied for spinal metastasis detection in MRI. In [31] a nucleus-guided feature extraction framework based on CNN was proposed for computer-aided diagnosis of breast cancer in histopathological images.

The rest of the paper is organized as follows. The problem formulation and the proposed context-sensitive DNN classifier are presented in Section 2. The experiments for evaluating the performance of the proposed DNN detector in detecting both individual MCs and MC clusters are descried in Section 3. The evaluation results are presented in Section 4. Finally, conclusions are given in Section 5.

2. Context-sensitive deep neural network model

2.1. Motivation and overview of MC detector architecture

In this study we formulate MC detection as a two-class classification problem, wherein a classifier is employed to determine whether an MC object is present (class 1) or absent (class 0) at a location under consideration in a mammogram

image. The classifier is trained through supervised learning, for which examples of MC objects and non-MC objects are used to optimize the classifier model.

As noted in the introduction, given the localized nature of individual MCs, it is desirable for the classifier to examine the image features only within a small neighborhood around an MC object. On the other hand, it is beneficial to also examine the image features in the surrounding background of an MC in order to suppress the occurrence of potential FPs in detection. A straightforward solution would be to simply apply the detector to an image window that is much larger in extent than the MC. However, this can be problematic, because the individual MCs in a cluster may occur in close vicinity of each other.

Out of the above considerations, in this study we propose a context-sensitive detector by taking into account both the local image features and the surrounding image background of an MC. As illustrated in Figure 2, at a detection location, we apply two image windows as input to the classifier simultaneously, one for characterizing the image features of an MC object while the other for describing the properties of the surrounding image background. For the classifier, we consider a deep neural network (DNN) structure. Accordingly, it consists of two subnetworks, one operating on the local image window and the other on the surrounding image background, as illustrated in Figure 3. Conceptually, the subnetwork in the second branch is used to extract the local image features of an MC object (hence termed local subnetwork); it plays a similar role to a local MC detector [6, 32]. The subnetwork in the first branch is used to characterize the contextual information surrounding an MC (hence termed global subnetwork), which is used to suppress potential FP objects in detection (such as those in linear structures). The resulting image features from the two branches are fed together into the fully-connected layers for classifying whether the input object is an MC or not. The two subnetworks are trained to learn and optimize the relevant contextual and local image features simultaneously. In image classification problems it was demonstrated that features extracted by deep learning could yield improved performance over hand-crafted features [33, 34].

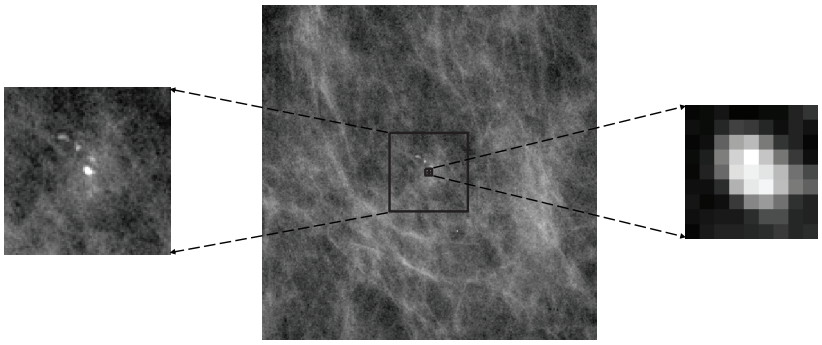


Figure 2: At a detection location, a small image window is used for characterizing the image features of an MC object, whereas a large image window is used for its surrounding tissue background.

2.2. Architecture of context-sensitive DNN

As illustrated in Figure 3, the input to the proposed DNN classifier is formed by two image windows co-centered at a detection location. For convenience, let $M_1 \times M_1$ and $M_2 \times M_2$ denote the sizes of these two windows ($M_1 > M_2$), respectively. In concept, the $M_1 \times M_1$ window is chosen to be sufficiently large so that it is representative of the image context around the detection location, whereas the $M_2 \times M_2$ window should be small enough to cover the extent of a single MC object (to avoid overlapping with other MCs).

The two subnetworks in the DNN classifier are each formed by a cascade of several convolutional (Conv), batch normalization, nonlinearity, and max-pooling (Pooling) layers, followed by several fully connected (FC) layers for final classification output. The composition of these layers for each subnetwork is optimized through model training (as to be described in Section 2.4). The details of these layers are as follows.

Convolutional layers are the core layers in the context-sensitive DNN classifier. They are used to extract the features in the input image at varying spatial scales. Each Conv layer produces a number of feature maps by convolving its input with a set of convolutional kernels. Mathematically, let $\mathbf{x}$ be the input, $\mathbf{w}_k$ be the kth convolutional kernel, and $\mathbf{y}_k$ be the corresponding output, $k = 1, 2, \cdots, K$, where K is the number of convolutional kernels. Then the

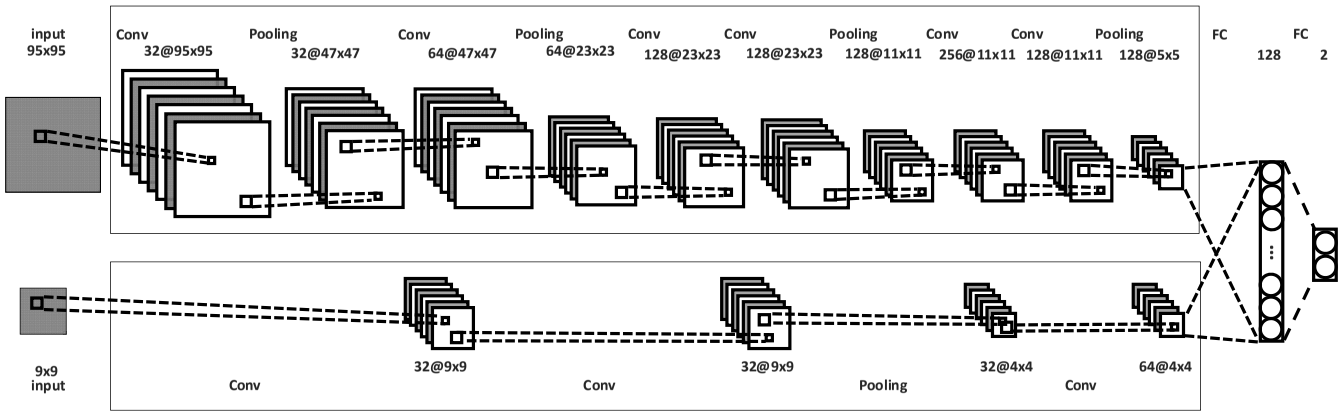


Figure 3: Illustration of the proposed context-sensitive DNN classifier architecture. It consists of two subnetworks, one for processing the large image context window (called global subnetwork), and one for processing the small MC image window (called local subnetwork). A batch normalization layer and a nonlinearity layer are included immediately after each Conv layer, which are not shown in the figure for brevity.

output of a Conv layer can be described by:

$$\mathbf{y}_k = \mathbf{x} * \mathbf{w}_k + b_k \tag{1}$$

where $*$ denotes the convolution operation and b_k is a bias. There are typically several (i.e. $K > 1$) convolutional kernels in each Conv layer. For example, as indicated in Figure 3, there are 32 convolutional kernels used in the first Conv layer of the global subnetwork, each producing a feature map of 95×95 pixels. In this study, all the convolutional kernels are set to be 3×3 in size.

Batch normalization layers are used to deal with the issue of internal covariate shift during training, a phenomenon that the distribution of a layer's input values varies with the parameters of its proceeding layers [35]. These normalization layers are used to independently normalize the feature values at each layer to zero mean and unit standard deviation during each training batch, wherein the normalization parameters are estimated from all the samples in the training batch. Such a normalization step can speed up learning and improve classification accuracy [35]. In the proposed context-sensitive DNN classifier, a batch normalization layer is applied to each Conv layer for feature map normalization; due to space limitation, these layers are omitted in Figure 3.

Nonlinearity layers are used to produce a non-linear transformation on the output of the corresponding neurons. In this study, we consider rectified linear units (ReLU) as follows:

$$f(x) = \max(0, x) \tag{2}$$

Compared with sigmoid activation functions, ReLU units can lead to faster training [36, 37] and yield sparse representations [37]. Due to space limitation, these layers are omitted in Figure 3 as well.

Max-pooling layers are used to achieve non-linear down-sampling of the feature maps [38]. The rationale is to enable the Conv layers to extract image features at increased scales. In this study, the output at each max-pooling layer is generated for every other location (i.e. with stride 2) by taking the maximum

value in the 3×3 neighborhood of the location.

Finally, two *fully-connected layers* are used in Figure 3. They play the same role as in a standard feed-forward neural network [36]. The two layers have 128 and 2 neurons, respectively. The input to the first fully connected layer is formed by all the features from the two subnetworks.

In the final output, a softmax activation function is used [39]. The output can be interpreted as the probability of an input belonging to one of the two classes.

2.3. Variants of DNN architecture for optimization

The particular architecture in Figure 3 was used to illustrate the context-sensitive DNN classifier. In this study, we investigated several variants of this DNN architecture for optimizing MC detection. These variants were obtained by varying the number of Conv layers used in the two subnetworks. Such choice was based on the observation that the number of Conv layers tends to have a larger impact on performance than other parameters (such as the number of filters and filter size) in the design of deep architectures [40].

Global subnetwork: The structure of the global subnetwork is motivated by the work previously developed in [18], where a CNN classifier was applied for determining whether a mammogram region contains clustered MCs (instead of a single MC) or not. We adopt this classifier here for characterizing the image background (i.e., context) surrounding a detection location. For this network, the size M_1 of the input image window was set to be 95 in the experiments; this corresponds to an image region of approximately 1 cm^2 (image resolution at 0.1 mm/pixel). Such a setting was out of the consideration that clustered MCs are typically localized within a 1 cm^2 area.

For optimizing the architecture, we varied the number of Conv layers increasingly from three to eight. The corresponding structures of these variants are summarized in Table 1. Note there is a batch normalization layer and a non-linearity layer immediately after each convolutional layer; they are not listed in Table 1 for brevity. A note is that, compared to the CNN classifier in [18], in

this study batch normalization layers are used in place of local response normalization (LRN) layers in order to avoid the potential need for optimizing their associated parameters.

Table 1: The different structures considered for the global subnetwork in this study. There is a batch normalization layer and a nonlinearity layer after each Conv layer, but not listed in this table for brevity. For each structure, the presence of a layer is marked by ✓sign.

	# filters	feature maps	3 Convs	4 Convs	5 Convs	6 Convs	7 Convs	8 Convs
Conv	32	32×95×95	✓	✓	✓	✓	✓	✓
Conv	32	32×95×95	-	-	-	-	-	✓
Pooling	-	32×47×47	✓	✓	✓	✓	✓	✓
Conv	64	64×47×74	✓	✓	✓	✓	✓	✓
Conv	64	64×47×47	-	-	-	-	✓	✓
Pooling	-	32×23×23	✓	✓	✓	✓	✓	✓
Conv	128	128×23×23	✓	✓	✓	✓	✓	✓
Conv	128	128×23×23	-	-	-	✓	✓	✓
Pooling	-	32×11×11	✓	✓	✓	✓	✓	✓
Conv	256	256×11×11	-	✓	✓	✓	✓	✓
Conv	128	256×11×11	-	-	✓	✓	✓	✓
Pooling	-	256×5×5	-	✓	✓	✓	✓	✓

Local subnetwork: The local subnetwork is intended to characterize the image features of an MC object. For this purpose, the input image window to this network needs to be comparable to that of an MC in size. In the experiments, the size M_2 of this window was set to be 9; such a choice was found to be adequate in the previous studies on MC detectors [6, 32].

As in the global subnetwork, we also optimized the structure of the local subnetwork. Considering that the input image window is much smaller in the latter, we varied the number of Conv layers from one to four. The corresponding structures of these variants are summarized in Table 2. As in the global subnetwork, there is a batch normalization layer and a nonlinearity layer after each convolutional layer; they are omitted in Table 2 for brevity.

In summary, there were four variants considered for the local subnetwork and six variants for the global subnetwork considered in this study. During model

Table 2: The different structures considered for the local subnetwork in this study. There is a batch normalization layer and a nonlinearity layer after each Conv layer, but omitted in this table for brevity. For each structure, the presence of a layer is marked by ✓sign.

	# filters	feature maps	1 Conv	2 Convs	3 Convs	4 Convs
Conv	32	32×9×9	✓	✓	✓	✓
Conv	32	32×9×9	-	-	✓	✓
Pooling	-	32×4×4	✓	✓	✓	✓
Conv	64	64×4×4	-	✓	✓	✓
Conv	64	64×4×4	-	-	-	✓

training, these variants were used in combinations for the two subnetworks, yielding a total of $4 \times 6 = 24$ different structures. The one with the best validation error was chosen in the end for testing the detection performance.

2.4. Model training and optimization

For a given network structure, the various associated parameters are determined during the training phase. This is accomplished by minimizing the binomial logistic loss on the set of training samples, or equivalently, the cross-entropy between the model output and the actual labels of training samples [41].

For model training, we extracted a large number of MC samples and non-MC samples from an allocated set of mammograms (Section 3.2). Similarly, for model selection, we also extracted a large of MC samples and non-MC samples for an allocated set of mammograms for validation (Section 3.1). These validation samples were independent of the training samples. They were used for computing the validation errors of different DNN structures after training.

We implemented the proposed context-sensitive DNN classifier using the Keras package with Theano backend. For model training, the adaptive moment estimation (Adam) method [42] was used with the following parameters: batch size $= 256$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. For model validation, a model structure was assessed on the validation set after every 1,000 iterations of training with the maximum number of iterations set at 20,000; here one iteration refers to an update of the model parameters by a batch of training

samples. In the end, the model structure with the smallest validation error was selected as the optimal model.

As to be described later in 3.2, there were far more non-MC samples than MC samples in the training set. To keep the samples from the two classes in a more balanced way, we adopted the following strategy in training: 1) for the first 10,000 iterations, we randomly selected 128 MC samples and 128 non-MC samples in each batch; 2) for the second 10,000 iterations, we simply randomly selected 256 samples from all the training samples in each batch for training. The second step was used for fine-tuning model training based on training set.

Finally, to overcome potential model overfitting, a stochastic dropout technique [43] was applied during training. This dropout procedure is a regularization technique in which different neural units and their connections are randomly dropped from the network with a certain probability (0.5 was used in our experiments).

3. Experiments

3.1. Mammogram dataset

In the experiments we demonstrated the proposed approach using both screen-film mammogram (SFM) images and full-field digital mammogram (FFDM) images. We made use of 521 SFM images from 297 cases (151 benign/146 cancer), and 188 FFDM images from 95 cases (52 benign/43 cancer). All these images were collected by the Department of Radiology at the University of Chicago. They were consecutive cases collected over different time periods, and were all sent for biopsy due to the subtlety of their MC lesions. Each mammogram image had at least one cluster of MCs which was histologically proven. The SFM images were acquired using a Lumiscan film digitizer (Lumisys; Sunnyvale, CA). The FFDM images were acquired using a Senographe 2000D FFDM system (General Electric Medical Systems; Milwaukee, WI). All the images were of 0.1 mm/pixel in spatial resolution. The MCs in each mammogram were manually identified by a group of experienced radiologists. In total, there are 27,022

MCs in these mammograms, which were used as ground truth in our evaluation.

The collection of mammogram images was randomly partitioned into three subsets, one with 167 cases (300 images) for training, one with 67 cases (117 images) for validation, and one with 158 cases (292 images) for testing. It is noted that most of the cases had multiple views (mediolateral oblique view, cranio-caudal view, or views from both breasts). To avoid any potential bias, the different views from one case were assigned together to either the training, validation, or testing subset exclusively.

Prior to MC detection, a background subtraction step was first applied to a mammogram image under consideration in order to suppress the inhomogeneity in the tissue background. For each location, the background was estimated as the average intensity of a circular region with a diameter of 7 pixels centered at the location [13]. Afterward, the resulting image was normalized to have zero mean and unit standard deviation.

3.2. Extraction of data samples for DNN training and model validation

From each mammogram in the training set, we first extracted the MC class samples from the marked MCs in the image. At each marked MC location, two image windows (as indicated earlier in Figure 2) were cropped to form one input sample to the classifier. To balance the number of samples from different mammograms, no more than 150 MC samples were extracted when a particular mammogram had a large number of MCs. In the end there were a total of 9,197 MC samples extracted.

Similarly, we extracted the non-MC class samples from the training mammograms. Given that the majority area in a mammogram does not have any MCs, to increase the number of training samples, we extracted k times as many non-MC samples as MC samples from each mammogram. These non-MC samples were obtained from locations randomly selected in the background tissue area (without any MCs) of the mammogram. We varied k increasingly as 2, 4, 8, 16, and 20, and determined that the validation error of the trained classifier plateaued at 20. Thus, $k = 20$ was used in the end.

To further enlarge the number of training samples, we applied a data augmentation procedure for improving the performance of the DNN classifier [36, 44]. Specifically, we applied the following operations to augment the extracted image samples in the training set: 1) flipping the image from left to right, 2) flipping the image up-down, and 3) rotating the image by 90°, 180°, and 270°, respectively. The number of training samples was increased by five times after data augmentation. Note that these operations did not alter the spatial resolution of the samples, which is important in detection of MCs in a mammogram.

Finally, for the purpose of model validation, we also extracted a set of MC samples and non-MC samples from the set of 117 mammograms allocated for model validation using the same procedure described above. However, no data augmentation was applied for these validation samples. There were a total of 4,739 MCs in these validation mammograms. These validation samples were used to determine the best DNN architecture obtained from training.

3.3. Test 1: Accuracy of detecting individual MCs in image regions

To demonstrate the performance of the proposed DNN classifier, we first evaluated its accuracy in detecting individual MCs. For this purpose, we allocated 125 mammograms from the test set. For each mammogram, two image ROIs were cropped, one containing clustered MCs, and one without any MCs. The ROI with MCs was set to be 500×500 or 1000×1000 pixels according to the size of the MC cluster; the ROI without any MCs was set to be 500×500 pixels. This resulted in a total of 125 ROIs with clustered MCs and 125 ROIs without. We assessed the accuracy of the detected MCs by the DNN classifier in these image regions.

To summarize the detection performance, we conducted a free-response receiver operating characteristic (FROC) analysis of the MC detection results. An FROC curve is a plot of the true-positive (TP) fraction of the MCs detected versus the average number of FPs per unit image region ($1cm^2$ in area) with the decision threshold varied over an operating range [45]. For clarity, this is referred to as *MC-based FROC analysis*.

In the MC-based FROC analysis, the TP fraction was computed from the average of the TP fractions of the ROIs with clustered MCs, whereas the FP rate was computed from both the ROIs with and without any MCs [46]. In the detector output, a detected object was treated as a TP when at least 40% of its area overlaps with that of a true MC or its distance to the center of a true MC is no larger than 0.3 mm; otherwise it was counted as an FP [45].

Given that most of the image area in a mammogram does not have any MCs, in the experiments we first applied a pre-detection procedure to speed up the detection process. Instead of applying the DNN detector to every location in the image, we first obtained a set of potential MC candidates by using the difference-of-Gaussians (DoG) detector [5]. The DNN classifier was then applied to identify the MCs from this set of candidates. The DoG detector was chosen because of its computational simplicity. It consists of two Gaussian kernels with width parameters set as $\sigma_1 = 1.1$ and $\sigma_2 = 1.4$ [5]. To insure high sensitivity in detection, the decision threshold was deliberately set to a low value such that it yielded a TP rate of 93.9% in the training images; at this level, it also yielded a high rate of 104.9 FPs per unit region (1 cm^2).

To accommodate the variations associated with case selection and facilitate statistical comparisons, we applied a bootstrapping procedure in the MC-based FROC analysis on the set of mammogram ROIs [47]. In our experiments a total of 20,000 bootstrap samples were used, based on which the partial area under the FROC curve (pAUC) was obtained.

3.4. Test 2: Accuracy of detecting MC clusters in mammograms

We also evaluated the performance of the proposed approach in detection of MC clusters from mammograms. In this test, our goal is to assess the accuracy when the DNN classifier is applied to detect the presence of MC clusters in a mammogram. For this purpose, we used the 167 images (113 SFM, 54 FFDM) from the remaining 84 cases in the test set.

To evaluate the detection performance, we conducted an FROC analysis on the detection results from the test mammograms. Different from the MC-based

FROC analysis above, the FROC analysis here is to study the accuracy of the detected MC clusters (rather than individual MCs) in the mammograms. To avoid any potential confusion, this FROC study is referred to as *cluster-based FROC analysis*. In this analysis, the FROC curve is a plot of the true-positive (TP) rate of detected MC clusters versus the average number FP clusters per image with the decision threshold varied continuously over its operating range.

In cluster-based FROC analysis, a detected MC cluster was considered as a TP according to the following criteria [48]: (1) it includes at least two true MCs; and (2) its center of gravity is within 1 cm of that of a known true MC cluster region. Likewise, a detected cluster was counted as an FP provided that (1) it contains no true MCs, or (2) the distance between its center of gravity and that of any known cluster region is larger than 1 cm. It is noted that the cluster-based FROC curve can be sensitive to the detection criteria used [48]. However, the relative performance by different detection methods with respect to a common set of criteria tends to be consistent [6, 9, 18].

To detect the MC clusters in a mammogram, the DNN classifier was applied to detect the presence of individual MCs first. Afterward, the detected MCs were grouped into clusters by using a connecting distance of 0.5 cm [18]. Those clusters with fewer than three detections were discarded. The remaining detected clusters were determined to be TP or FP according to the criteria described above.

To reduce the effect of case variations, we applied a bootstrapping procedure on the set of test mammograms for obtaining the cluster-based FROC [47]. A total of 20,000 bootstrap samples were used, based on which pAUC was obtained. To speed up the FROC analysis, in the experiments we first applied a pre-scouting step as in [9], during which up to four most suspicious regions of $5cm \times 5cm$ in size were identified in each mammogram image for further consideration. Afterward, the proposed DNN classifier was applied for detecting the MCs in these regions (as in Test 1 above).

3.5. Methods for comparison

To demonstrate the overall performance of the proposed context-sensitive DNN classifier, we also considered the following MC detection methods in the experiments:

1) Unified SVM detector: Recently developed in [9], this MC detector was designed to suppress occurrence of FPs caused by linear structures and MC-like noise patterns. It was demonstrated to yield improved performance over several existing MC detectors [9].

2) CNN cluster detector: This detector was developed for direct detection of the presence of clustered MCs in an image region based on deep learning [18]. It was demonstrated to outperform several existing MC based detectors in detecting MC clusters.

3) A local DNN detector: In contrast to the proposed context-sensitive DNN classifier, we also considered a DNN detector by using only the local subnetwork in the DNN architecture (Figure 3). This is to demonstrate the benefit of the additional use of the global subnetwork for context learning in MC detection. For convenience, this detector is referred to as local DNN. In the experiment, the local DNN classifier was trained with the same set of training samples by using the same procedure as in the context-sensitive DNN.

4. Results and Discussions

4.1. Model selection and model robustness

In Table 3 we summarize the validation results obtained by the proposed context-sensitive DNN classifier with various model structures. In this table, the classification errors on the MC and non-MC samples in the validation set (Section 3.2) are given for different variants of the global and local subnetworks. As can be seen, the best classification error of 0.533% was obtained by the DNN classifier with three Conv layers in the local subnetwork and seven Conv layers in the global subnetwork (Figure 3). Consequently, this model structure was selected for subsequent performance evaluation.

Table 3: Validation errors achieved by the context-sensitive DNN classifier with different architectures of the local and global subnetworks.

Local subnetwork	Global subnetwork					
	3 Convs	4 Convs	5 Convs	6 Convs	7 Convs	8 Convs
1 Conv	0.739%	0.655%	0.638%	0.682%	0.646%	0.766%
2 Convs	0.657%	0.633%	0.554%	0.569%	0.579%	0.560%
3 Convs	0.725%	0.581%	0.656%	0.627%	**0.533%**	0.571%
4 Convs	0.672%	0.556%	0.644%	0.545%	0.671%	0.603%

It is noted from Table 3 that the validation errors are consistently small for the different structures, with the worst case being 0.766% obtained with one Conv layer in the local subnetwork and eight Conv layers in the global subnetwork. These results indicate that the DNN classifier performance is relatively robust to the variations in the two subnetworks. In addition, in Figure 4 we show a plot of the training and validation errors vs. the number of iterations obtained by the optimal architecture; in the end, the one with the smallest validation error (e.g., 9,000 iterations) was selected.

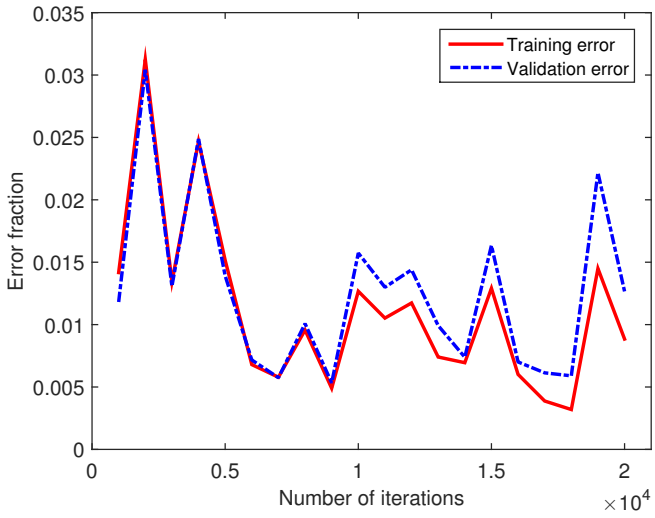


Figure 4: Plot of training and validation errors for every 1,000 iterations obtained by the context-sensitive DNN architecture with three Conv layers in the local subnetwork and seven Conv layers in in the global subnetwork.

4.2. Test 1: Accuracy of detecting individual MCs in image regions

In Figure 5 we show the MC-based FROC curve obtained by the context-sensitive DNN classifier in detecting individual MCs on the set of test mammogram ROIs (Section 3.1). In this plot, the y-axis represents the fraction of true MCs detected (i.e., sensitivity) by the classifier, whereas the x-axis represents the number of FPs detected per unit mammogram area (1 cm^2). For comparison, the MC-based FROC curve obtained by the Unified SVM detector is also shown in Figure 5. As can be seen, the FROC curve of the context-sensitive DNN classifier is notably higher (hence better detection performance). A statistical comparison between the two classifiers yielded a difference of 1.15 in pAUC (p-value $< 10^{-4}$; 95% C. I. $= [0.97, 1.34]$) for FP rate over the range of [0, 10]. In particular, with TPF at 80%, the context-sensitive DNN classifier achieved an FP rate of 1.03 FPs/cm^2, compared to 5.69 FPs/cm^2 by the Unified SVM (a reduction of 81.9%). Moreover, with FP rate at 1 FP/cm^2, the context-sensitive DNN classifier achieved a sensitivity of 79.7%, compared to 66.1% for the Unified SVM.

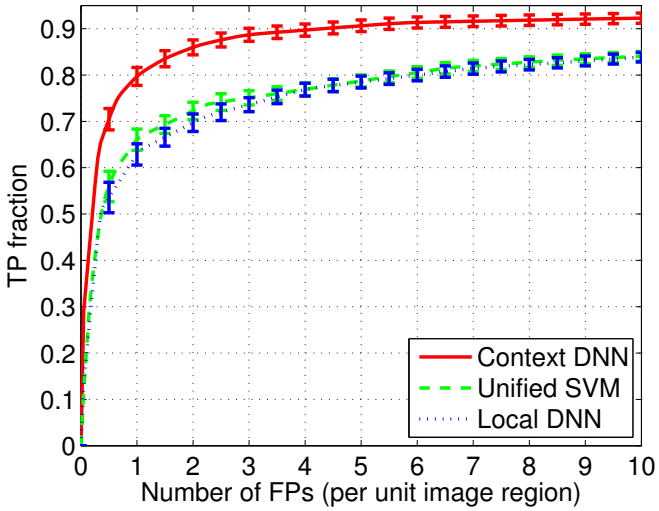


Figure 5: MC-based FROC curves obtained by different classifiers in detecting individual MCs: 1) context-sensitive DNN (Context DNN), 2) Unified SVM, and 3) local DNN.

Moreover, to demonstrate the effect of context learning by the global sub-network in the DNN classifier, in Figure 5 we also show the MC-based FROC curve obtained by the local DNN; this classifier has three Conv layers as the context-sensitive DNN classifier. As can be seen, the FROC curve of the local DNN is much lower compared to the context-sensitive DNN. A statistical comparison between the two yielded a difference of 1.26 in pAUC (p-value $< 10^{-4}$; 95% C. I. $= [1.07, 1.46]$) for FP rate over the range of [0, 10] FPs/cm^2. In particular, with TPF at 80%, the context-sensitive DNN classifier achieved an FP rate of 1.03 FPs/cm^2, compared to 6.00 FPs/cm^2 by the local DNN (a reduction of 82.8%). Moreover, with FP rate at 1 FPs/cm^2, the context-sensitive DNN classifier achieved a sensitivity of 79.7%, compared to 62.9% for the local DNN. These results clearly indicate the benefit of context learning for improving the detection accuracy of individual MCs.

4.3. Test 2: Accuracy of detecting MC clusters in mammograms

In Figure 6 we show the cluster-based FROC curve obtained by the context-sensitive DNN classifier in detection of MC clusters in mammograms. In this plot, the y-axis represents the fraction of true MC clusters detected (i.e., sensitivity) among the test mammograms, while the x-axis represents the average number of FP clusters per image. For comparison, the cluster-based FROC curve obtained by the Unified SVM is also shown. As can be seen, the FROC curve is higher for the context-sensitive DNN classifier; a statistical comparison between the two yielded a difference of 0.10 in pAUC (p-value $= 0.0060$; 95% C. I. $= [0.02, 0.18]$) for FP rate over the range of [0, 2] clusters/image. In particular, with TPF at 85%, the context-sensitive DNN classifier achieved an FP rate of 0.40 clusters/image, compared to 0.52 clusters/image by the Unified SVM classifier (a reduction of 22.2%). Moreover, with FP rate at 0.5 clusters/image, the optimal context-sensitive DNN classifier achieved a sensitivity of 87.4%, compared to 84.6% for the Unified SVM classifier.

In addition, in Figure 6 we also show the cluster-based FROC curve obtained by the local DNN classifier. As can be seen, the FROC curve is much lower than

the context-sensitive DNN; a statistical comparison between the two yielded a difference of 0.15 in pAUC (p-value < 0.0001; 95% C. I. $= [0.076, 0.24]$) for FP rate over the range of [0, 2]. In particular, with TPF at 85%, the context-sensitive DNN classifier achieved an FP rate of 0.40 clusters/image, compared to 1.10 clusters/image by the local DNN classifier (a reduction of 63.5%). Moreover, with FP rate at 0.5 clusters/image, the context-sensitive DNN classifier achieved a sensitivity of 87.4%, compared to 76.4% for the local DNN classifier. These results indicate that the context learning in the context-sensitive DNN is beneficial for improving the detection accuracy of MC clusters.

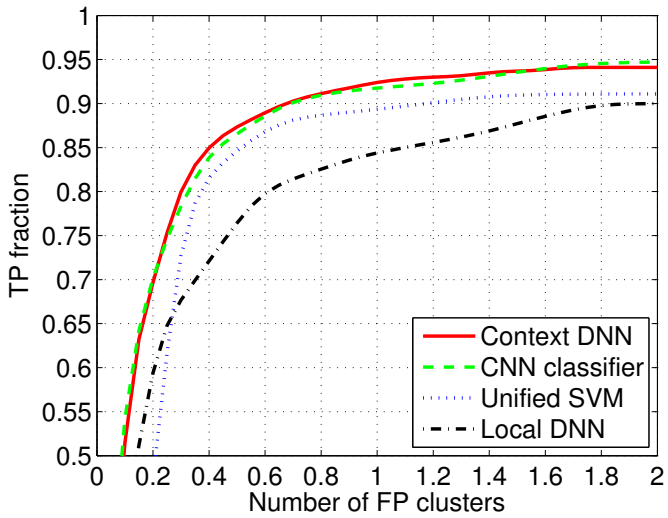


Figure 6: Cluster-based FROC curves obtained by different classifiers in detecting MC clusters: 1) context-sensitive DNN (Context DNN), 2) Unified SVM, 3) local DNN, and 4) CNN cluster detector.

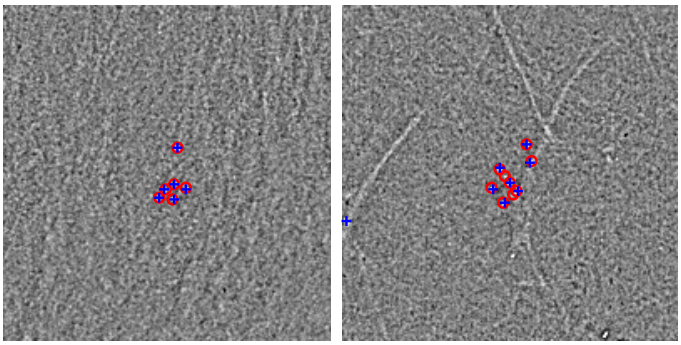
Finally, in Figure 6 we also show the FROC curve obtained by the cluster-based CNN classifier in [18]; this classifier was for direct detection of MC clusters rather than individual MCs. Interestingly, the FROC curve of the CNN classifier is noted to be very close to the context-sensitive DNN classifier; a statistical comparison between the two yielded a difference of 0.0032 in pAUC (p-value $= 0.4699$; 95% C. I. $= [-0.058, 0.068]$) for FP rate over the range of [0, 2] clusters/image. In particular, with TPF at 85%, the context-

sensitive DNN classifier achieved an FP rate of 0.40 clusters/image, compared to 0.44 clusters/image by the CNN classifier. Moreover, with FP rate at 0.5 clusters/image, the context-sensitive DNN classifier achieved a sensitivity of 87.4%, compared to 86.6% for the CNN classifier. These results indicate that the two classifiers can achieve essentially similar accuracy in detection of the presence of MC clusters in mammograms. However, compared to the CNN classifier, the proposed context-sensitive DNN classifier is able to provide the locations of individual MCs for further analysis.

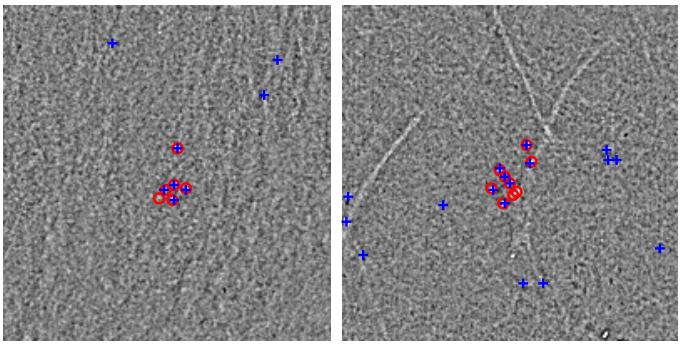
In the experiments, the proposed context-sensitive DNN classifier was implemented in a GPU card of GeForce GTX TITAN X with 12 GB memory. For the selected model structure, it took about four and a half hours in training, and 12.18 seconds per mammogram on average for detecting MC clusters during testing.

4.4. Detection examples

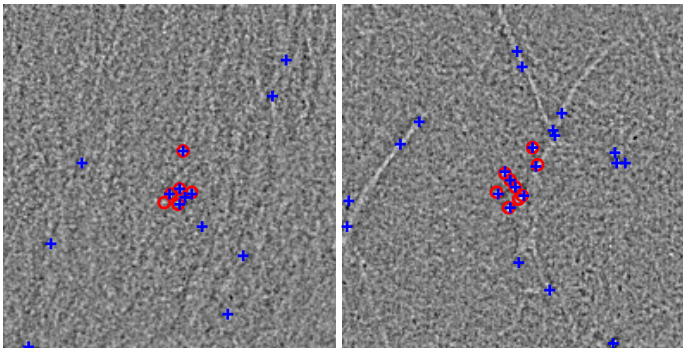
For illustration, in Figure 7 we show the detection results obtained by the different detectors for the two ROIs shown earlier in Figure 1. In these results, the operating points were set for the different detectors such that the TP rate was 70% in detecting individual MCs. As can be seen, while most of the MCs were correctly detected by the different detectors, there were fewer FPs in the context-sensitive DNN classifier than the Unified SVM and the local DNN. Specifically, for the first ROI (left), all the MCs were correctly detected without any FPs in the context-sensitive DNN, compared to one MC missed and three FPs in the Unified SVM and one MC missed and eight FPs in the local DNN. For the second ROI (right), there were two MCs missed and one FP in the context-sensitive DNN, compared to two MCs missed and 10 FPs in the Unified SVM and one MC missed and 15 FPs in the local DNN. Note that two FPs in the Unified SVM and nine FPs in the local DNN were associated with linear structures.



(a) Context-sensitive DNN



(b) Unified SVM



(c) Local DNN

Figure 7: Detection results achieved by different classifiers for the two example ROIs in Figure 1. In each ROI, the MCs are indicated by red circles, and the detections are indicated by blue plus signs.

4.5. Discussions

The evaluation results above demonstrate that incorporating global context information into the proposed DNN classifier can lead to improvement in the detection accuracy of both individual MCs and MC clusters in mammograms. MC detectors are developed traditionally for detecting the presence of clustered MCs in mammograms which are intended as an alert for examination. Consequently, their detection performance is typically evaluated on the accuracy of the detected clusters. However, the accuracy of detected individual MCs is also important in CADx applications where a detected cluster is further analyzed for classification as being benign or malignant. Therefore, we employed both MC-based FROC and cluster-based FROC for measuring the detection performance of the proposed approach.

From the cluster-based FROC results (Test 2) it is seen that the context-sensitive DNN detector could achieve a similar level of accuracy in detected MC clusters to that the CNN classifier which was designed for direct MC cluster detection. This is remarkable in that MC based detectors are more likely susceptible to FPs caused by MC-like noise and image patterns. This was evident from the FROC results obtained by the local DNN classifier. Moreover, the global subnetwork employed in the context-sensitive DNN classifier is similar to the CNN classifier for same global context processing. It is reasonable that the context-sensitive DNN could match the results by CNN in terms of detected MC clusters. More importantly, the proposed DNN could provide the important result of individual MC locations.

In this study, the input image window sizes (i.e. M_1 and M_2) for the two subnetworks were chosen based on the typical physical size of individual MCs and the spatial extent of MC clusters, as explained earlier in Section 2. However, in the future, it might be interesting to further investigate what would be the optimal combination of the two input windows. In addition, we considered CNN structures for both local and global subnetworks in this study. It would be also interesting to investigate the use of the alternative architectural structures in the proposed context-sensitive detector framework, such as ResNet [49], Inception

[39, 50], and squeeze-and-excitation network [51]).

5. Conclusion

In this study, we developed a context-sensitive DNN classifier for detecting clustered MCs in a mammogram. The classifier takes into account both the local image features and the surrounding image background in its decision at a detection location. It consists of two subnetworks, one for extracting the local image features of an MC object, and the other for characterizing its surrounding background. We evaluated the proposed approach using FROC analysis on a set of 292 screen-film and full-field digital mammogram images. The FROC analysis was conducted both on the accuracy of detected individual MCs and the MC clusters. The results show that the context-sensitive DNN could achieve significantly more accurate results in detecting both individual MCs and MC clusters when compared to a local DNN classifier and the Unified SVM detector; moreover, it could also match the accuracy of a global CNN classifier in detecting MC clusters. Specifically, with TPF at 80%, the context-sensitive DNN achieved an FP rate of 1.03 FPs/cm^2, compared to 5.69 FPs/cm^2 for the Unified SVM and 6.00 FPs/cm^2 for the local DNN in detecting individual MCs. Similarly, with TPF at 85%, the context-sensitive DNN achieved an FP rate of 0.40 clusters/image, compared to 0.52 clusters/image for the Unified SVM and 1.10 clusters/image for the local DNN classifier in detecting MC clusters. These results indicate that incorporating image context information in MC detection can indeed improve the accuracy of detected MCs. Encouraged by these promising results, we plan to further investigate whether this classifier could also lead to improved performance in classifying a detected cluster as being benign or malignant.

References

[1] A. C. Society, Cancer facts and figures, Atlanta, GA, 2017.

[2] M. Lanyi, Diagnosis and differential diagnosis of breast calcifications, Springer-Verlag, 1988.

[3] D. Mintzer, J. Glassburn, B. A. Mason, D. Sataloff, Breast cancer in the very young patient: a multidisciplinary case presentation, The Oncologist 7 (6) (2002) 547–554.

[4] L. Wei, Y. Yang, R. M. Nishikawa, Y. Jiang, A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications, IEEE Transactions on Medical Imaging 24 (3) (2005) 371–380.

[5] M. F. Salfity, R. M. Nishikawa, Y. Jiang, J. Papaioannou, The use of a priori information in the detection of mammographic microcalcifications to improve their classification, Medical Physics 30 (5) (2003) 823–831.

[6] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, R. M. Nishikawa, A support vector machine approach for detection of microcalcifications, IEEE Transactions on Medical Imaging 21 (12) (2002) 1552–1563.

[7] H.-D. Cheng, J. Wang, X. Shi, Microcalcification detection using fuzzy logic and scale space approaches, Pattern Recognition 37 (2) (2004) 363–375.

[8] A. Oliver, A. Torrent, X. Lladó, M. Tortajada, L. Tortajada, M. Sentís, J. Freixenet, R. Zwiggelaar, Automatic microcalcification and cluster detection for digital and digitised mammograms, Knowledge-Based Systems 28 (2012) 68–75.

[9] J. Wang, R. M. Nishikawa, Y. Yang, Improving the accuracy in detection of clustered microcalcifications with a context-sensitive classification model, Medical Physics 43 (1) (2016) 159–170.

[10] J.-J. Mordang, T. Janssen, A. Bria, T. Kooi, A. Gubern-Mérida, N. Karssemeijer, Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks, in: International Workshop on Digital Mammography, Springer, 2016, pp. 35–42.

[11] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, K. Cha, M. A. Helvie, Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis, in: SPIE Medical Imaging, International Society for Optics and Photonics, 2016, pp. 97850Y–97850Y.

[12] J. Wang, Y. Yang, R. M. Nishikawa, Reduction of false positive detection in clustered microcalcifications, in: Image Processing (ICIP), 2013 20th IEEE International Conference on, IEEE, 2013, pp. 1433–1437.

[13] W. J. Veldkamp, N. Karssemeijer, Normalization of local contrast in mammograms, IEEE Transactions on Medical Imaging 19 (7) (2000) 731–738.

[14] K. J. McLoughlin, P. J. Bones, N. Karssemeijer, Noise equalization for detection of microcalcification clusters in direct digital mammogram images, IEEE Transactions on Medical Imaging 23 (3) (2004) 313–320.

[15] H.-P. Chan, S. Galhotra, C. J. Vyborny, H. MacMahon, P. M. Jokich, et al., Image feature analysis and computer-aided diagnosis in digital radiography. i. automated detection of microcalcifications in mammography, Medical Physics 14 (4) (1987) 538–548.

[16] A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, D. Romani, An SVM classifier to separate false signals from microcalcifications in digital mammograms, Physics in Medicine and Biology 46 (6) (2001) 1651.

[17] S. Chen, H. Zhao, False-positive reduction using RANSAC in mammography microcalcification detection, in: SPIE Medical Imaging, International Society for Optics and Photonics, 2011, pp. 79631V–79631V.

[18] J. Wang, R. M. Nishikawa, Y. Yang, Global detection approach for clustered microcalcifications in mammograms using a deep learning network, Journal of Medical Imaging 4 (2) (2017) 024501–024501.

[19] H. Soltanian-Zadeh, F. Rafiee-Rad, et al., Comparison of multiwavelet, wavelet, haralick, and shape features for microcalcification classification in mammograms, Pattern Recognition 37 (10) (2004) 1973–1986.

[20] B. Verma, P. McLeod, A. Klevansky, A novel soft cluster neural network for the classification of suspicious areas in digital mammograms, Pattern Recognition 42 (9) (2009) 1845–1852.

[21] I. Andreadis, G. Spyrou, K. Nikita, A comparative study of image features for classification of breast microcalcifications, Measurement Science and Technology 22 (11) (2011) 114005.

[22] Z. Chen, H. Strange, A. Oliver, E. R. Denton, C. Boggis, R. Zwiggelaar, Topological modeling and classification of mammographic microcalcification clusters, IEEE Transactions on Biomedical Engineering 62 (4) (2015) 1203–1214.

[23] J. Wang, Y. Yang, Boosted classification of breast cancer by retrieval of cases having similar disease likelihood, in: Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on, IEEE, 2016, pp. 908–911.

[24] Y. Jiang, R. M. Nishikawa, J. Papaioannou, Dependence of computer classification of clustered microcalcifications on the correct detection of microcalcifications, Medical Physics 28 (9) (2001) 1949–1957.

[25] J. Wang, Y. Yang, Spatial density modeling for discirminating between benign and malignant microcalcification lesions, in: Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on, IEEE, 2013, pp. 133–136.

[26] M. V. S. de Cea, R. M. Nishikawa, Y. Yang, Estimating the accuracy level among individual detections in clustered microcalcifications, IEEE Transactions on Medical Imaging 36 (5) (2017) 1162–1171.

[27] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA 316 (22) (2016) 2402–2410.

[28] R. Rasti, M. Teshnehlab, S. L. Phung, Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks, Pattern Recognition 72 (2017) 381–390.

[29] Z. Jiao, X. Gao, Y. Wang, J. Li, A parasitic metric learning net for breast mass classification based on mammography, Pattern Recognition 75 (2017) 292–301.

[30] J. Wang, Z. Fang, N. Lang, H. Yuan, M.-Y. Su, P. Baldi, A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks, Computers in Biology and Medicine 84 (2017) 137–146.

[31] Y. Zheng, Z. Jiang, F. Xie, H. Zhang, Y. Ma, H. Shi, Y. Zhao, Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification, Pattern Recognition 71 (2017) 14–25.

[32] L. Wei, Y. Yang, R. M. Nishikawa, M. N. Wernick, A. Edwards, Relevance vector machine for automatic detection of clustered microcalcifications, IEEE Transactions on Medical Imaging 24 (10) (2005) 1278–1285.

[33] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (8) (2013) 1915–1929.

[34] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, Overfeat: Integrated recognition, localization and detection using convolutional networks, arXiv preprint arXiv:1312.6229.

[35] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167.

[36] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[37] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks., in: Aistats, Vol. 15, 2011, p. 275.

[38] B. Graham, Fractional max-pooling, arXiv preprint arXiv:1412.6071.

[39] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[40] K. He, J. Sun, Convolutional neural networks at constrained time cost, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5353–5360.

[41] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv:1207.0580.

[42] D. Kingma, J. Ba, Adam: A method for stochastic optimization, in: International Conference on Learning Representations, 2015, pp. 1–13.

[43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, The Journal of Machine Learning Research 15 (1) (2014) 1929–1958.

[44] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the devil in the details: Delving deep into convolutional nets, arXiv preprint arXiv:1405.3531.

[45] J. Wang, R. M. Nishikawa, Y. Yang, Quantitative comparison of clustered microcalcifications in for-presentation and for-processing mammograms in full-field digital mammography, Medical Physics 44 (2017) 3726–3738.

[46] J. Wang, H. Ding, F. Azamian, B. Zhou, C. Iribarren, S. Molloi, P. Baldi, Detecting cardiovascular disease from mammograms with deep learning, IEEE Transactions on Medical Imaging 36 (5) (2017) 1172–1181.

[47] F. W. Samuelson, N. Petrick, Comparing image detection algorithms using resampling, in: Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on, IEEE, 2006, pp. 1312–1315.

[48] R. M. Nishikawa, Current status and future directions of computer-aided diagnosis in mammography, Computerized Medical Imaging and Graphics 31 (4) (2007) 224–235.

[49] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[50] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[51] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, arXiv preprint arXiv:1709.01507.