# CONTEXT-SENSITIVE DEEP LEARNING FOR DETECTION OF CLUSTERED MICROCALCIFICATIONS IN MAMMOGRAMS

*Juan Wang* and *Yongyi Yang*

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) for MC detection, aimed to take into account both the local image features of an MC and its surrounding tissue background. The proposed approach was evaluated on the accuracy both in detecting individual MCs and in detecting MC clusters on a set of 292 mammograms using free-response receiver operating characteristic (FROC) analysis. The results demonstrate that the proposed approach could achieve a significantly higher accuracy in detected individual MCs; incorporating image context information in MC detection can be beneficial for reducing FPs.

***Index Terms*—**Computer-aided diagnosis (CAD), clustered microcalcifications, 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 breast cancer deaths will occur among women in the US in 2017 [1]. Mammography is an effective screening tool for diagnosis of breast cancer. It can detect about 80%-90% of breast cancer cases in women without symptoms [1]. One important early sign of breast cancer is the appearance of clustered micro-calcifications (MCs), which are tiny calcium deposits that exhibit as bright spots in mammograms (Fig. 1). Clustered MCs can occur both in benign cases and in malignant cases.

In the literature, computerized methods for detection of clustered MCs are referred to as computer-aided detection (CADe), the purpose of which is to serve as an alert to the radiologists by detecting the presence of suspicious regions for examination in screening mammography. In these methods, detection of clustered MCs is typically performed in two steps [2-5]. 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.

While successful in achieving high sensitivity, a challenging issue facing MC detection methods is the frequent occurrence of FPs. This is because the response in an MC detector 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, inhomogeneity in tissue background, imaging artifacts, etc [6]. 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 surrounding tissues. For example, noise equalization techniques were developed for reducing the noise variability in a mammogram [7]; background removal methods were used to suppress the inhomogeneity in tissue background [8]; image features associated with linear structures were also incorporated for reducing FPs [4,9].

Recently, we developed a one-step approach [10] 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 (~1 $cm^2$) which contains not only the individual MCs but also their surrounding image context. This approach was demonstrated to be more robust to the FP patterns in mammograms when compared to several existing MC detectors [10]. However, this detector is intended only for identifying suspicious regions for subsequent examination; it does not specify the locations of individual MCs within a detected region.

Built on the success with our approach in [10], in this study we investigate the feasibility of extending it 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 [11-12]. Studies have shown that the accuracy of detected individual MCs can impact on the CADx performance [13-15].

Toward this goal, we develop a deep neural network (DNN) architecture in order to take into account not only the local image features of an MC but also its surrounding image context for MC detection. The detector network is formed by two subnetworks for classifying whether an MC is present or not at a detection location, one for extracting the local image features and the other for learning the image features of its surrounding background. Consequently, the detector response is automatically adapted to the image

background at an MC. 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.

## 2. METHODS

### 2.1 Motivation and overview of detector architecture

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

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 also beneficial to examine the image features in the surrounding background of an MC in order to suppress the occurrence of potential FPs. 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 may occur in close vicinity of each other.

Out of these 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 Fig. 2, at a detection location, we apply two co-centered 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, as illustrated in Fig. 3. It consists of two subnetworks, one operating on the local image window and the other on the surrounding image background. The resulting image features from the two subnetworks are fed together into the fully-connected layers for classifying whether the input object is an MC or not.

### 2.2 Architecture of context-sensitive DNN

As illustrated earlier in Fig. 3, the subnetwork in the first branch of the detector network is used to characterize the contextual information surrounding an MC (hence termed global subnetwork), whereas the subnetwork in the second branch is used to extract the local image features of an MC object (hence termed local subnetwork). The two subnetworks are trained to extract and optimize the relevant contextual and local image features simultaneously.

Conceptually, the input image window to the global subnetwork should be sufficiently large in extent so that it is representative of the image context surrounding the detection location (such as linear structures). In contrast, the input image window to the local subnetwork should be small enough to cover a single MC object (to avoid overlapping with other MCs), as in a local MC detector [3]. Based on these considerations, in the experiments, the global image widow was set as 95×95 pixels in size as in the direct MC cluster detector [10], and the local image window was set to be 9×9 pixels as in an MC detector [3].

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. In each subnetwork, Conv layers are used to extract the features in the input image at varying spatial scales. In this study, all the Conv kernels are set to be $3\times3$ in size. The batch normalization layers are used to deal with the issue of internal covariate shift during training [16]. The nonlinearity layers are used to produce a non-linear transformation on the output of the neurons; in this study, we consider rectified linear units (ReLU). The max-pooling layers are used to achieve non-linear down-sampling of the feature maps. The output at each max-pooling layer is generated for every other location (i.e. stride 2) by taking the maximum value in the $3\times3$ neighborhood of the location. The FC layers play the same role as in a standard feed-forward neural network. The input to the first FC layer is formed by all the features from the two subnetworks. In the final output, a softmax activation function is used. The output can be interpreted as the probability of an input belonging to one of the two classes.

### 2.3 Model training and validation

For model training and validation of the DNN, samples of MC objects were extracted for each of the marked MCs in a set of training mammograms (Sect. 3.1). For each MC object, both an MC window and a global image window were simultaneously extracted (Fig. 2) and were formed as one input. Given that there are no MCs in most of the area in a mammogram, 20 times as many non-MC samples as the MCs were also randomly extracted from the background tissue area of each mammogram. To balance the number of samples from different mammograms, no more than 150 MC samples were extracted from a mammogram.

To enlarge the set of training samples, the following augmentation operations were applied to each training sample: 1) flipping the image windows left-right, 2) flipping the image windows up-down, and 3) rotating the image windows by 90°, 180°, and 270°, respectively. No data augmentation was used for the samples in the validation set.

For network training, the various unknown parameters were determined by using the adaptive moment estimation (Adam) method [17] to minimize the binary entropy loss. To overcome potential model overfitting, a stochastic dropout technique [18] with probability 0.5 was applied during training. We experimented with many variants of the architecture by varying the number of Conv layers in the two subnetworks (3 to 8 for the global subnetwork, and 1 to 4 for the local subnetwork); the one with the lowest classification

error in the validation set is shown in Fig. 3 (7 and 3 Conv layers in the global and local subnetworks, respectively).

## 3. EXPERIMENTS AND RESULTS

### 3.1 Mammogram dataset

In this study we made use of 521 screen-film mammogram (SFM) images from 297 cases and 188 full-field digital mammogram (FFDM) images from 95 cases. All these images were collected by the Department of Radiology at the University of Chicago. Each mammogram was of 0.1 mm/pixel in spatial resolution and had at least one cluster of MCs which was histologically proven. The MCs in each mammogram were manually identified by a group of experienced radiologists, which were used as ground truth in our evaluation. The dataset 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.

Prior to MC detection, a background subtraction step as in [10] was applied to each mammogram in order to suppress the inhomogeneity in the tissue background. Afterward, the resulting image was normalized to have zero mean and unit standard deviation.

### 3.2 Performance evaluation

*3.2.1 Accuracy in detecting individual MCs*

We first evaluated the accuracy of the proposed DNN in detecting individual MCs from mammogram regions. For this purpose, we allocated 125 mammograms from the test set and cropped two regions (500×500 or 1000×1000 pixels depending on the cluster size) from each image, one containing clustered MCs, and one without any MCs. We then applied the proposed DNN detector to these two ROIs for MC objects.

To summarize the detection performance, we conducted a free-response receiver operating characteristic (FROC) analysis. 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 (1 $cm^2$ in area) with the decision threshold varied over an operating range [19]. For clarity, this is referred to as *MC-based FROC analysis*.

To speed up the detection process, we first used the difference-of-Gaussians (DoG) detector [2] to obtain a set of potential candidates of MCs in each image region, and then applied the proposed DNN classifier to these candidates to determine the MC objects.

*3.2.2 Accuracy in detecting MC clusters*

We also evaluated the performance of the proposed approach in detection of MC clusters from mammograms. For this purpose, we used the remaining 167 test images.

To evaluate the detection performance, we conducted a *cluster-based FROC analysis* on the accuracy of the detected MC clusters (rather than individual MCs) in the mammograms. In this analysis, the FROC curve is a plot of the TP rate of detected MC clusters versus the average number FP clusters per image.

To accommodate the variations associated with case selection and facilitate statistical comparisons, we applied a bootstrapping procedure in the (MC-based or clustered-based) FROC analysis [20]. 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.

### 3.3 Methods for comparison

We considered the following MC detection methods in the experiments: 1) proposed context-sensitive DNN detector, 2) unified SVM detector [3], 3) CNN cluster detector [10], and 4) a local DNN detector, formed by the local subnetwork in Fig. 3, which is used to demonstrate the benefit of additional context learning in MC detection.

### 3.4 Results

*3.4.1 Detecting individual MCs in image regions*

In Fig. 4 we show the MC-based FROC curve obtained by the proposed context-sensitive DNN classifier in detecting individual MCs on the set of test mammogram regions. For comparison, results are also given for the Unified SVM detector. The FROC curve of the DNN is notably higher (better detection performance). A statistical comparison between the two yielded a difference of 1.15 in $pAUC$ ($p$-value$<10^{-4}$) for FP rate over the range of [0, 10] FPs/$cm^2$.

Moreover, in Fig. 4 we also show the results obtained by the local DNN. As can be seen, the FROC curve of the local DNN is much lower compared to the context-sensitive DNN (difference in $pAUC$ = 1.26; $p$-value$<10^{-4}$). These results clearly indicate the benefit of context learning for improving the detection accuracy of individual MCs.

*3.4.2 Detecting MC clusters in mammograms*

In Fig. 5 we show the cluster-based FROC curves obtained by the proposed DNN and the Unified SVM in detection of MC clusters in mammograms. The FROC curve is higher for the DNN classifier; a statistical comparison between the two yielded a difference of 0.10 in $pAUC$ ($p$-value=0.006) for FP rate over the range of [0, 2] clusters/image.

For comparison, we also show in Fig. 5 the results obtained by the CNN cluster detector in [10] and the local DNN classifier. As can be seen, the local DNN yielded a much lower FROC curve than the context-sensitive DNN (difference in $pAUC$=0.15; $p$-value < 0.0001). Interestingly, the FROC curve of the CNN detector is noted to be very close to the context-sensitive DNN (difference in $pAUC$ = 0.0032; $p$-value = 0.4699).

The FROC results of the local DNN detector clearly indicate the vulnerability in detection due to FPs caused by local image patterns resembling MCs. As noted in the introduction, the CNN cluster detector was designed to directly detect MC clusters in mammograms without specifically identifying the locations of individual MCs. Interestingly, by utilizing both local image features and

global context information, the proposed DNN detector could significantly improve the accuracy in detecting individual MCs while without sacrificing the accuracy in detecting MC clusters achieved by the CNN detector.

## 4. CONCLUSION

We developed a context-sensitive DNN classifier for detecting clustered MCs in mammograms. 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 detector using FROC analysis on a set of 292 screen-film and full-field digital mammogram images. The results indicate that incorporating image context information in MC detection can significantly improve the accuracy in detecting individual MCs.

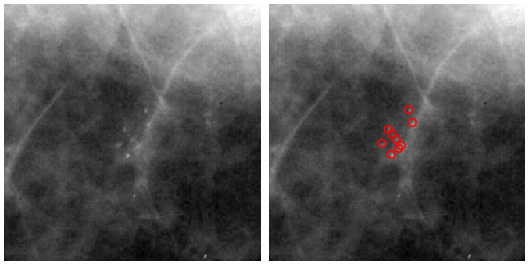

**Fig. 1** *Left:* Example ROI (200×200 pixels, 0.1 mm/pixel) containing clustered MCs; *Right:* Locations of individual MCs marked by red circles.

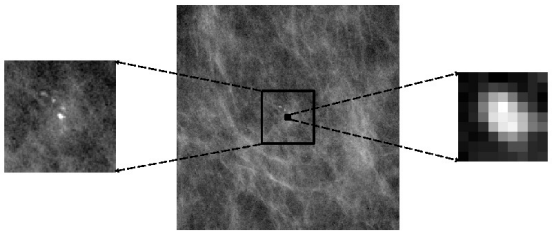

**Fig. 2** At a detection location, a small image window is used for characterizing the image features of an MC object, whereas a large image window is used for its surrounding tissue background.

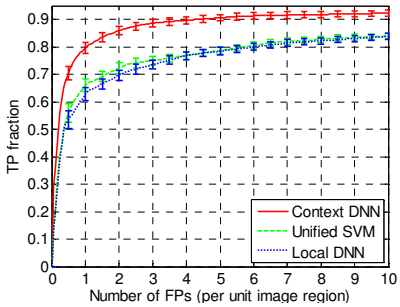


**Fig. 4** MC-based FROC curves obtained by different detectors in detecting individual MCs: 1) context-sensitive DNN (Context DNN), 2) Unified SVM, and 3) local DNN.

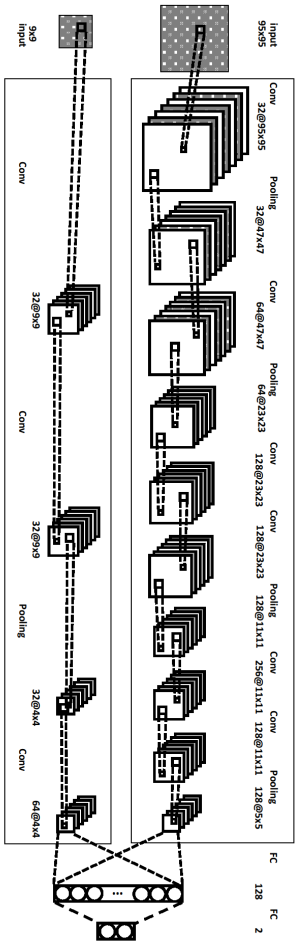


**Fig. 3** Illustration of the proposed context-sensitive DNN classifier architecture. It consists of two subnetworks, one for processing the large image context window (called global subnetwork), and one for processing the small MC image window (called local subnetwork). A batch normalization layer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

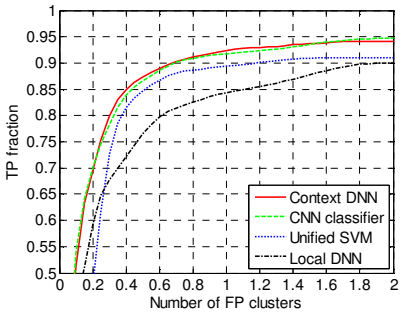


**Fig. 5** Cluster-based FROC curves obtained by different detectors in detecting MC clusters: 1) context-sensitive DNN (Context DNN), 2) Unified SVM, 3) local DNN, and 4) CNN cluster detector.

## 5. REFERENCES


[1] American Cancer Society, "Cancer facts and figures", Atlanta, GA, 2017.

[2] M. F. Salfity, R. M. Nishikawa, Y. Jiang, and J. Papaioannou, "The use of a priori information in the detection of mammographic microcalcifications to improve their classification," *Medical Physics*, vol. 30, no. 5, pp. 823–831, 2003.

[3] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.

[4] J. Wang, R. M. Nishikawa, and Y. Yang, "Improving the accuracy in detection of clustered microcalcifications with a context-sensitive classification model," *Medical Physics*, vol. 43, no. 1, pp. 159–170, 2016.

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

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

[7] K. J. McLoughlin, P. J. Bones, and N. Karssemeijer, "Noise equalization for detection of microcalcification clusters in direct digital mammogram images," *IEEE Transactions on Medical Imaging*, vol. 23, no. 3, pp. 313–320, 2004.

[8] H.-P. Chan, S. Galhotra, et al., "Image feature analysis and computer-aided diagnosis in digital radiography. i. automated detection of microcalcifications in mammography," *Medical Physics*, vol. 14 no. 4, pp. 538–548, 1987.

[9] S. Chen and H. Zhao, "False-positive reduction using RANSAC in mammography microcalcification detection," *SPIE Medical Imaging*, pp. 79631V–79631V, 2011.

[10] J. Wang, R. M. Nishikawa, and Y. Yang, "Global detection approach for clustered microcalcifications in mammograms using a deep learning network," *Journal of Medical Imaging*, vol. 4, no. 2, pp. 024501–024501, 2017.

[11] H. Soltanian-Zadeh, F. Rafiee-Rad, et al., "Comparison of multiwavelet, wavelet, haralick, and shape features for microcalcification classification in mammograms," *Pattern Recognition*, vol. 37, no. 10, 1973–1986, 2004.

[12] Z. Chen, H. Strange, et al., "Topological modeling and classification of mammographic microcalcification clusters," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1203–1214, 2015.

[13] Y. Jiang, R. M. Nishikawa, and J. Papaioannou, "Dependence of computer classification of clustered microcalcifications on the correct detection of microcalcifications," *Medical Physics*, vol. 28, no. 9, pp. 1949–1957, 2001.

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

[15] M. V. S. de Cea, R. M. Nishikawa, and Y. Yang, "Estimating the accuracy level among individual detections in clustered microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 36, no. 5, pp. 1162–1171, 2017.

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

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

[18] 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*, vol. 15, no. 1, pp. 1929–1958, 2014.

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

[20] F. W. Samuelson and N. Petrick, "Comparing image detection algorithms using resampling," *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pp. 1312–1315, 2006.