Sādhanā (2020)45:44
https://doi.org/10.1007/s12046-019-1260-4

Reduction of false positives in the screening CAD tool for microcalcification detection

VIKRANT A KARALE[1], TULIKA SINGH[2], ANUP SADHU[3], NIRANJAN KHANDELWAL[2] and SUDIPTA MUKHOPADHYAY[1,*]

[1]Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, India
[2]Department of Radiodiagnosis and Imaging, Post-graduate Institute of Medical Education and Research, Chandigarh 160012, India
[3]EKO CT & MRI Scan Center, Kolkata Medical College, Kolkata 700004, India
e-mail: smukho@ece.iitkgp.ac.in

MS received 23 October 2019; accepted 18 November 2019

Abstract. Breast cancer is one of the leading causes of cancer deaths among women worldwide. Early diagnosis of breast cancer can help in reducing the mortality rate. The major challenge in the early diagnosis of breast cancer is the fewer number of radiologists available per million population in developing countries. The total number of radiologists is less than 30 in many third world countries. Since majority of the screening mammograms are normal or do not show any cancer signs, there is need of a screening computer-aided diagnosis (CAD) tool that can detect normal mammograms correctly and thereby reduce the burden on radiologists. Thus, a screening CAD is developed that is able to detect microcalcification clusters in mammogram with 100% sensitivity on the subset of DDSM, INbreast and PGIMER-IITKGP databases at lower false positives as compared with state of the art methods. The synthetic minority over-sampling technique and the majority class under-sampling based on data distribution are used to improve the classifiers performance by reducing the false positives. An approach based on principal component analysis is proposed to further reduce the false positives by removing the vascular calcifications that are not of any clinical significance and may increase the false positives.

Keywords. 2D-NEO; nonlinear energy operator; HOG; CAD; breast cancer; microcalcifications.

1. Introduction

Breast cancer is the second most frequent type of cancer diagnosed and the major factor behind cancer deaths in the women worldwide [1]. Early diagnosis of breast cancer could be helpful in the reduction of mortality rate, cost of the treatment and the anxiety of the patients. Breast cancer can be diagnosed in the non-palpable stage with the help of mammography [2]. It has been found that mammography screening has resulted in 19% reduction of mortality rates caused by the breast cancer [3]. American Cancer Society (ACS) recommends women to undertake screening mammography, starting at the age of 45 years [4]. There are various indicators of breast cancer present in the mammograms like mass, architectural distortion, microcalcification (MC) clusters and bilateral asymmetry. The MC clusters can be the only early indicator of breast cancer and they are also reported in almost 93% cases of ductal carcinoma in situ [2, 5]. Hence, MC cluster detection is vital in the breast cancer prognosis.

There is massive workload on the radiologist due to fewer number of radiologists available in most of the countries. The number of radiologists per million population is 40.7, 42.1 and 52.1 in Japan, Germany and Australia, respectively. Some of the African countries have one radiologist for 1 million population and a few of them have less than 30 radiologists [6]. The majority of screening mammograms are normal or do not contain any signs of cancer. Hence, there is need of a screening computer-aided diagnosis (CAD) tool that can detect normal mammograms accurately and minimize the burden of the radiologists.

Several studies have been conducted on MC detection in mammograms. Many local and global preprocessing techniques have been proposed to increase contrast between MC and other high-intensity regions. Papadopoulos *et al* [7] have investigated various preprocessing methods such as wavelet-based preprocessing, contrast-limited adaptive histogram equalization and linear range modification. The

*For correspondence
Published online: 10 February 2020

LRM-based preprocessing achieved the best results. Kim and Park [8] proposed the features extracted using surrounding region-dependence method (SRDM) for the MC detection. Soltanian-Zadeh *et al* [9] have shown that the multi-wavelet-based features perform better than shape, texture and wavelet-based features for MC detection.

Yu and Guan [10] proposed various features such as grey level statistics, wavelet features and shape features. The two neural networks are trained using these features and a sensitivity of 90% is reported at 0.5 false positives (FPs) per image. A successive enhancement learning is proposed by El-Naqa *et al* [11] to train a support vector machine (SVM) using imbalanced data. Wei *et al* [12] proposed relevance vector machines for locating MC clusters and obtained the sensitivity of 90% with one FP per image. Peng *et al* [13] proposed an algorithm to detect MC clusters based on stochastic resonance noise. Nakayama *et al* [14] proposed a multi-scale features extracted from a filter bank based on Hessian matrix to distinguish between nodular and the linear structures in mammograms. Oliver *et al* [15] proposed bank of filters to extract local features and used boosting classifier to detect MC clusters. Zhang *et al* [16] useda technique based on the top hat transform and wavelet transform to detect MC clusters. The technique obtained 92.9% sensitivity at 0.08 FP per image. Shin *et al* [17] proposed a Discriminative Restricted Boltzmann Machine, which discriminates between normal tissues and MCs. The method reported A_z value of 0.8294. Liu *et al* [18] applied possibilistic fuzzy *c*-mean clustering to cluster MCs and weighted SVM for the detection of MC clusters. The method obtained 92% sensitivity at 2.3 FP per image. Guo *et al* [19] used pulse-coupled neural network and contourlet transform for the detection of MC clusters. Mordang *et al* [20] proposed a hard negative mining strategy to handle class imbalance problem. Convolutional neural network (CNN) is trained to classify manually extracted patches of 13×13 pixels to detect MC clusters. A sensitivity of 99.92% is obtained at FP rate of 0.1. Karale *et al* [21] proposed a modified unsharp masking to suppress the background and enhance the MC pixels. The method achieved 96.72% sensitivity with 3.48 FP per image and 96.05% sensitivity with 1.81 FP per image on Digital Database for Screening Mammography (DDSM) and private database, respectively.

Wang *et al* [22] used a CNN-based approach to locate MC clusters and reported 90% sensitivity at 0.69 FP per image on a private database. Wang and Yang [23] proposed context-sensitive deep neural network and reported 85% sensitivity with 0.4 FP per image on a private database. Most of the approaches mentioned in the literature do not automatically localize the MC clusters and classify manually selected regions as MC clusters or normal regions. Karale *et al* [24] proposed two techniques using mean multi-scale 2D NEO (MnM2DNEO) and max multi-scale 2D NEO (MxM2DNEO), which can act as screening tools to reduce workload of radiologists by detecting considerable portion of total normal mammograms. In this work, the two techniques by Karale *et al* [24] are improved by incorporating synthetic minority over-sampling technique (SMOTE) to handle data imbalance effectively, and also a novel method is proposed to discard vascular calcification, which results in reduction of FP clusters.

2. Methodology

2.1 *Database*

Three different databases, including scanned film and Full-Field Digital Mammography (FFDM), are used to assess the performance of proposed techniques. The scanned film database includes DDSM [25], which is a popular, publicly available database. The FFDM databases include INbreast [26] database, which is available publicly, and PGIMER-IITKGP database, which is acquired from PGIMER, Chandigarh, India. The mammograms of DDSM database have a pixel dimension of 42, 43.5 and 50 μm, and have a bit depth of 16 bits. In this study, 100 mammograms without MC clusters and 97 mammogram having 100 MC clusters are selected. Subtlety ratings in the range of 1–5 are provided by the DDSM database. The subtlety rating of 1 indicates the cases that are most difficult to detect, whereas 5 indicates the easiest to detect cases. For the unbiased assessment of proposed techniques, 20 MC clusters from each subtlety level are selected in this study. The PGIMER-IITKGP database has a pixel dimension of 70 μm and a bit depth of 12 bits. A set of 110 mammograms, including 50 mammograms without MC clusters and 60 mammograms having a minimum of one and maximum of five MC clusters, are chosen for this study. The INbreast database [26] has a pixel dimension of 70 μm and a bit depth of 14 bits. It has a total of 410 mammograms, which includes 389 mammograms without MC clusters and 21 mammograms having a minimum of one and maximum of three MC clusters. The information about the cluster boundaries is made available by the DDSM and INbreast databases, while the MCs are annotated by the expert radiologist AS. In case of PGIMER-IITKGP database, MCs are annotated by the expert radiologist TS. The INbreast database reports 21 mammograms with 27 MC clusters. Since it is noticed that one of the clusters contains less than three MCs, radiologist AS has combined that cluster with the neighbouring cluster. Thus, we have 26 MC clusters in 21 mammograms from INbreast database. The information about the three databases used in this study is given in table 1.

2.2 *Proposed method for MC cluster detection*

2.2a *Breast region segmentation:* The mammograms contain around 30% of the breast region. There can be many artefacts like labels or tags outside breast region in the

Table 1. Details of the database used in the study.

Database	Total no. of images	No. of images with MC	No. of MC clusters	No. of images without MC clusters
DDSM	197	97	100	100
INbreast	410	21	26	389
PGIMER-IITKGP	110	60	76	50

mammogram, which may be falsely detected as MCs. Thus, breast region segmentation can help in discarding these artifacts and can greatly reduce the computations by limiting the search space within breast region. Breast region is segmented using multilevel hierarchical thresholding (MLHT) [27, 28].

2.2b *Enhancement of MCs:* MCs are tiny deposits of calcium and appear as bright spots in mammograms. Detection of MCs is a challenge because of their subtle nature, irregular shapes and variable sizes. The two techniques proposed by Karale *et al* [24], namely MnM2DNEO and MxM2DNEO, are applied to increase the contrast of MC pixels over the background in mammograms.

The MxM2DNEO response at location (x, y) is defined by

$$\psi_1(x,y) = \max_{s\in 1,2,\ldots,S_m} 4sI^2(x,y) - \left[\sum_{v=-s}^{s} I(x-s,y+v)\times I(x+s,y-v) + \sum_{h=-(s-1)}^{s-1} I(x-h,y-s)\times I(x+h,y+s)\right] \quad (1)$$

The MnM2DNEO response at location (x,y) is defined as

$$\psi_2(x,y) = \frac{1}{S_m}\sum_{s=1}^{S_m} 4sI^2(x,y) - \left[\sum_{v=-s}^{s} I(x-s,y+v)\times I(x+s,y-v) + \sum_{h=-(s-1)}^{s-1} I(x-h,y-s)\times I(x+h,y+s)\right] \quad (2)$$

where I is the input image; h and v represent, respectively, shift in x and y; s represents the scale and S_m in the equations 2 and 1 is the maximum scale at which response of 2D NEO is computed. The value of S_m is calculated as follows:

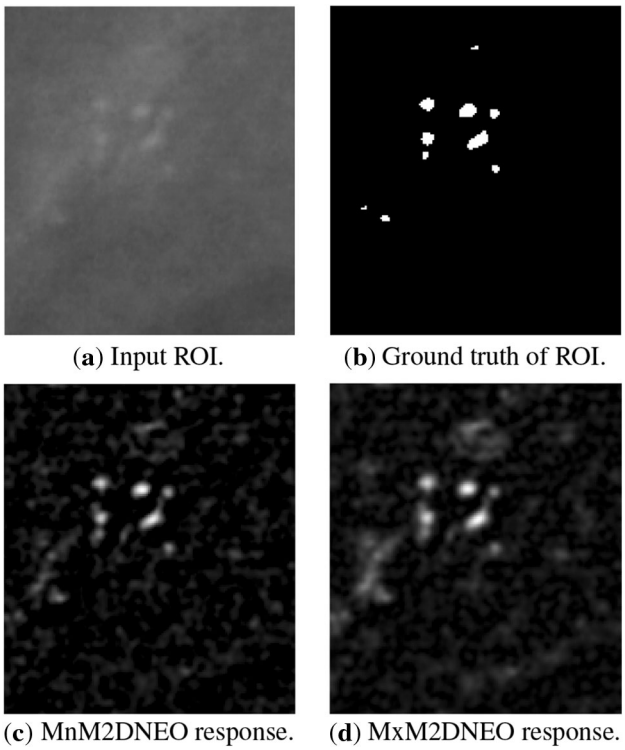
$$S_m = floor\left(\frac{D_m}{2}\right) \quad (3)$$

where D_m is the maximum diameter of the MC in pixels. Since the maximum size of the MC is 1 mm [11, 29], the value of D_m is estimated as

$$D_m = 2\left(round\left(\frac{1}{2r}\right)\right) + 1 \quad (4)$$

where r is the pixel dimension of the mammogram in millimeters. Since D_m represents maximum size of the window within which the 2D NEO is applied, the value of D_m is always maintained odd using Eq. (4). Figure 1(c) and (d) shows the response of MnM2DNEO and MxM2DNEO on input ROI, cropped from the image of DDSM database, respectively. The pre-processing step enhances the MC pixels and suppresses the other brighter regions in the mammogram.

2.2c *Iterative thresholding:* After the enhancement of MCs using multi-scale 2D NEO, the enhanced image is



(**a**) Input ROI. (**b**) Ground truth of ROI.
(**c**) MnM2DNEO response. (**d**) MxM2DNEO response.

Figure 1. The input ROI and its ground truth are shown in (a) and (b), respectively. The response of MnM2DNEO and MxM2DNEO for the input ROI is shown in (c) and (d), respectively.

thresholded to get the potential candidates of MCs. The thresholding step is implemented such that all the MC clusters are detected at the cost of higher FPs, which can be eliminated in the succeeding steps of the proposed algorithm.

For the estimation of threshold, the enhanced image is binarized at various threshold levels by varying the threshold from maximum value (T_{max}) to minimum value (T_{min}) in small steps (Δ_T). The T_{max} and T_{min} are the maximum and minimum intensities of the enhanced image, respectively, and Δ_T represents small change in threshold levels, given by $\Delta_T = (T_{max} - T_{min})/N$. The number of threshold levels is N. The number of detected objects is counted from the binary image, obtained at each threshold level, using connected component labelling. The threshold is decreased at each iteration till the number of detected objects is greater or equal to the preset number.

If the MCs are very close to each other, then they may get merged as a single detected object while thresholding the enhanced image. The merging of nearby MCs results in detected objects with the size bigger than size of MCs, which affects their morphological properties. In order to avoid this, the enhanced image is thresholded using multiple local thresholds instead of a single global threshold. Let the threshold at n^{th} level of thresholding be represented by t_n. If at n^{th} level of thresholding, multiple neighbouring objects (detected at $(n-1)^{th}$ level) get connected to form a single object then those objects are thresholded with the previous higher threshold t_{n-1} while other objects are thresholded with lower threshold t_n. This process of iterative thresholding is continued till the count of detected objects is above the preset number of MC candidates. The procedure to obtain the value of preset number is given in Karale *et al* [24].

2.2d *FP reduction using morphological features of MCs:* MCs are tiny bright spots in mammograms whose size varies from 0.05 to 1 mm [11]. Thus, the FPs that have major axis length greater than 1 mm can be rejected. In addition, small objects whose area is less than 3 pixels are detected due to the noisy pixels in the image. Thus, the detected objects that satisfy at least one of the following criteria are eliminated:

1. $l_{obj} > 1$ mm,
2. $A_{obj} < 3$ pixels.

2.2e *Feature extraction:* The FPs whose area is less than 3 pixels or whose diameter is less than 1 mm are discarded in the morphology-based FP reduction step. There are still many FP objects detected whose size is in the same range as that of MCs. Thus, the local features based on intensity, shape and texture are used to discriminate the MCs from the FP objects. For the purpose of extracting local features, the centroid of the individual MC is identified using connected component analysis. The local features are extracted from a square window whose centre is positioned at the centroid of the individual MC. The window size is equal to the major axis length of the detected object plus 6 pixels. A total of 38 local features [24] are computed from the enhanced image, which are given in table 2.

2.2f *Feature selection:* In this work, minimum redundancy maximum relevance (mRMR) [34] is used to choose most relevant features that can distinguish between the MCs and the FP objects. The features that have highest correlation with the target class and least redundancy with other features are selected. A combined score of relevance with target class and redundancy with selected feature is computed for each feature and the features are ranked accordingly. The incremental search is performed on the features using SVM classifier. The top k features that give the highest balanced accuracy [35] are selected.

2.2g *Addressing imbalance in training data:* The number of MCs is significantly lower compared with the detected FP objects. In the training data, the ratios of positive class samples (MCs) to negative class samples (FP objects) are about 1:90, 1:80 and 1:20 for the INbreast, DDSM and PGIMER-IITKGP databases, respectively. Machine learning algorithms are trained by minimizing the error on the training samples. Since there is less probability of occurrence of positive class samples, the cost of misclassifying the positive class samples is minimal. Thus, the classifier tends to get biased towards the majority class (negative class) data.

The class imbalance can be reduced using under-sampling technique, which decreases the count of samples present in the majority class. If the count of samples is inadequate to represent the minority class, then over-sampling technique can help in improving the classifier performance. Thus, SMOTE [36] is applied first to increase the number of minority samples. The minority class samples are increased by inserting samples between the selected minority data sample and each of its K nearest neighbours. For the estimation of value of K, 2-fold cross-validation is performed on the training set. The value of K is chosen such that it results in maximum mean balanced accuracy of the classifier.

After applying SMOTE, the under-sampling of majority class data can help in further reducing the data imbalance. The selection of a subset of majority class data is very crucial for the effective training of the classifier. The classifier can be trained effectively by selecting the hard samples or the samples that are nearer to the decision boundary [11]. Thus, a majority class data reduction technique [24] is used to reduce the count of majority class samples (negative class samples). The technique randomly samples the majority data all of whose features lie in the overlapping region of the distribution of positive and

Table 2. List of all the extracted features.

Sl. no.	Name	Description	Expression	Dimension
1	Mean intensity of foreground	Average intensity of foreground region within local window	$\mu_f = \sum_{i \in S_f} x(i)/\lvert S_f \rvert$ where S_f is set of foreground pixels in local window and $\lvert S_f \rvert$ is cardinality of set S_f	1
2	Standard deviation of foreground	Standard deviation of foreground region within local window	$\mu_f = \sqrt{\sum_{i \in S_f} (x(i) - \mu_f)^2/\lvert S_f \rvert}$	1
3	Relative foreground–background difference	Relative difference between mean foreground intensity and mean background intensity within local window	$FBR = (\mu_f - \mu_b)/(\mu_f + \mu_b)$ where μ_f and μ_b are mean intensities of foreground and background, respectively	1
4	Foreground–background difference	Difference between mean foreground intensity and mean background intensity within local window	$FBD = \mu_f - \mu_b$	1
5	Foreground entropy	Entropy of the foreground intensities within the local window	$H_f = -\sum_i p_i \log_2(p_i)$	1
6	Area	Count of pixels present in the detected object	$Area = \lvert S_f \rvert$	1
7	Compactness	Compactness of the detected object	$Compactness = area/P^2$ where P is number of pixels in contour of object	1
8	Shape moments	Shape moments represent the shape properties of detected object	The shape moments are computed as per Shen *et al* [30]	3
9	Elongation	Elongation is the ratio of major to minor axis length of ellipse fitted on the pixels of detected object	$Elongation = \frac{maj\_axis\_len}{min\_axis\_len}$ where $maj\_axis\_len = 4\lambda_1$, $min\_axis\_len = 4\lambda_2$. Here, λ_1 and λ_2 are eigenvalues of fitted ellipse	1
10	Invariant moments	The invariant moments are invariant to rotation, translation and scale of the input image	Seven invariant moments are calculated from local window as given in the Gonzalez and Woods [31] book	7
11	Haralick features	Thirteen texture-based features [32] are extracted within the local window containing the object	Texture features are extracted using 8×8 GLCM matrix computed with $d = 1$ and $\theta = 0°$	13
12	HOG-based features	HOG obtained from the local window is quantized into 4 bins. The magnitudes of 4-bin HOG are used as features. Also, mean, standard deviation and kurtosis are computed from HOG.	The HOG features [33] are computed from the local window containing the detected object. The histogram of 4 bins corresponding to 0°, 90°, 180°, 270°are computed to extract shape information of the objects. Three statistical features like mean, standard deviation and kurtosis values are also computed using the histogram.	4 + 3

negative data. The overlapping region for a 'foreground entropy' feature from DDSM database is shown in figure 2. The count of samples selected from the majority class for training is equal to the count of minority class samples.

2.2h *Classification:* SVM [37] classifier is used to discriminate between the true MCs and the other detected objects. The radial basis function is used as kernel function in the SVM. The regularization parameter (C) and the Gaussian width (σ) of the radial basis function is estimated using the grid search. The training set of 5-fold validation is used to perform grid search. At each iteration of 5-fold cross-validation, the training set is further divided using 2-fold cross-validation. The values of C and σ that yield maximum mean cross-validation accuracy are selected.

2.2i *Nearest neighbour clustering:* MC clusters are often the only early indicator of breast cancer [5]. On the other hand, isolated MCs do not have clinical significance. Thus, the proposed algorithm is concentrated on the MC cluster detection. A cluster of MC consists of three or more MCs in 1 cm^2 region [38]. Hence, a nearest neighbour clustering with maximum distance of 0.5 cm is applied to group the detected objects into clusters. The clusters having less than three MCs are rejected.

2.2j *Removal of vascular calcification:* The vascular calcifications are always benign and not clinically significant for diagnosing breast cancer [39]. These calcifications are calcium deposits deposited on the walls of the blood vessels. They are often detected by MC detection algorithms, which increases false alarms. Thus, a novel and simple algorithm is proposed to discard vascular calcifications.

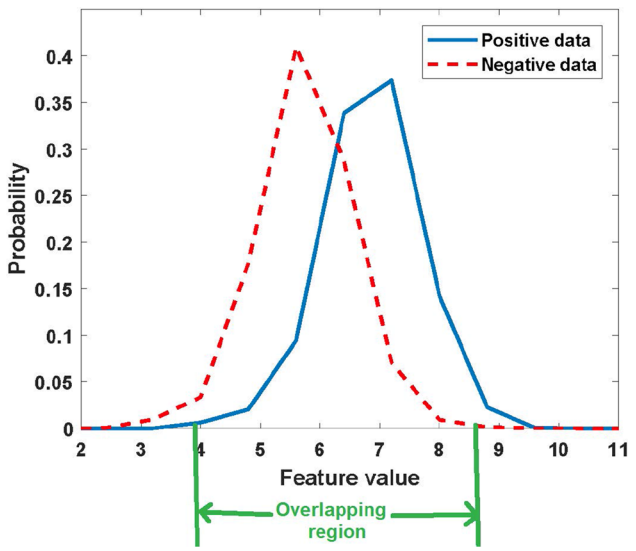


Figure 2. Probability distribution of positive and negative data of feature 'Entropy from Haralick features' from DDSM database.

The vascular calcifications and MC cluster are shown in figure 3(a) and (c), respectively. From each detected cluster, the centroids of all the detected objects are computed using connected component labelling. The principal component analysis (PCA) is applied on these centroids for each cluster. In case of vascular calcification cluster, the difference between the magnitude of the eigenvectors is very high as compared with the case of MC cluster (without vascular calcification) as indicated in figure 3(b) and (d). The eigenvalues give variance or spread of the detected objects along the two eigenvectors, and the sum of eigenvalues gives total variance. The percentage of total variance (PTV) explained by principle component is used to measure the difference between the variance along the eigenvectors. Higher value of PTV explained by principle component indicates higher differences between the variances along eigenvectors. The percentage of total variance explained by principle component is defined as

$$PTV_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2} \tag{5}$$

where λ_1 and λ_2 are the eigenvalues obtained from PCA and $\lambda_1 > \lambda_2$. Thus, the clusters, having PTV explained by principal component greater than predefined threshold, are discarded.

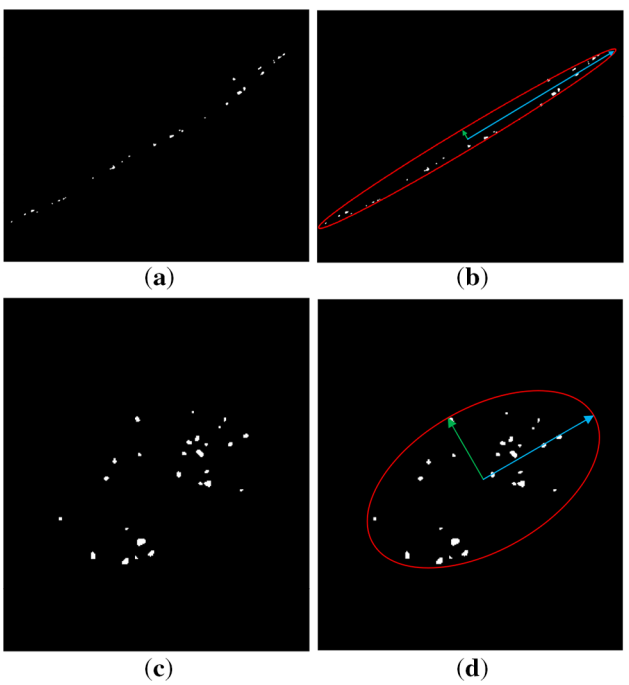


Figure 3. (a) Detected vascular calcification. (b) Ellipsoid (red colour) fitted on the vascular calcification; major and minor axes are shown in blue colour. (c) Detected MC cluster. (d) Ellipsoid (red colour) fitted on the MC cluster; major and minor axes are shown in blue colour.

3. Results and discussion

The results of the proposed techniques are compared to those of the complete automated techniques such as SVM-SEL by El-Naqa *et al* [11], top hat transform by Zhang *et al* [16], unsharp masking by Karale *et al* [21], MnM2DNEO by Karale *et al* [24] and MxM2DNEO by Karale *et al* [24]. The results of the competing techniques are simulated on DDSM, INbreast and PGIMER-IITKGP database for bench-marking.

Five-fold cross-validation is performed for each database to assess the performance of proposed and competing techniques. Stratified sampling is done on the sets of mammograms with and without MC clusters for each database. Also the data is split into five folds such that all the mammograms of the same patient are either present in training or testing set. The parameters and features are estimated from the training data of five-fold validation. DDSM database recommends taking into account the distribution of subtlety ratings while presenting the results on the subset of database. Thus, we have selected the mammogram images such that they have equal number of MC clusters of each subtlety level in the test set.

The MnM2DNEO and MxM2DNEO use under-sampling proposed by Karale *et al* [24]. Since the count of positive samples is low, under-sampling selects a few negative samples to maintain ratio of positive to negative samples. Thus, the classifier may not get trained effectively with insufficient amount of data in both classes. In case of modified MnM2DNEO and modified MxM2DNEO, SMOTE is used to synthetically generate the number of positive samples and then under-sampling based on data distribution is used to maintain equal number of samples in positive and negative class. Thus, the classifier is trained on increased number of samples, which improves the performance of the classifier. Also the vascular calcification removal stage in modified MnM2DNEO and modified MxM2DNEO further reduces the FPs. Table 3 shows the effect of applying SMOTE and vascular calcification removal step with respect to sensitivity and FP per image.

For the comparison of performance, the operating point is selected such that maximum sensitivity is achieved by a technique for each fold of five-fold cross-validation. For the purpose of analysing the robustness of each technique, the mean and standard deviation of the maximum sensitivity, and the corresponding mean and standard deviation of FP per image, are computed across five folds. The comparative performance with respect to mean maximum sensitivity with the corresponding mean and standard deviation of FP per image for proposed and competing techniques is presented in table 4. The modified mean multi-scale NEO has the lowest deviation in FP per image in case of the DDSM and INbreast databases, while unsharp masking has the minimum deviation in case of the PGIMER-IITKGP database. The modified mean multi-scale NEO has nearly same deviation in FP per image as obtained for unsharp masking, in case of PGIMER-IITKGP database. Thus, the modified mean multi-scale NEO is comparatively the most robust and best performing technique with nearly same deviation in FP per image.

Figures 4, 5 and 6 show the FROC plots of proposed techniques with competing techniques, namely MnM2DNEO [24], MxM2DNEO [24], modified unsharp masking [21], top hat transform [16] and SVM-SEL [11] for INbreast, DDSM and PGIMER-IITKGP database, respectively. From figure 4, MxM2DNEO has the best performance for lower sensitivities, but it could not achieve 100% sensitivity for INbreast database. The modified MnM2DNEO reached sensitivity of 100% at lower FP per image compared with MnM2DNEO. As shown in figure 5, modified MxM2DNEO performs better compared with MxM2DNEO. The MnM2DNEO has less FP per image compared with modified MnM2DNEO at lower sensitivities. The modified MnM2DNEO reached sensitivity of 100% at less FP per image compared with MnM2DNEO. In case of PGIMER-IITKGP database, the modified MnM2DNEO has the best performance, followed by

Table 3. Results of modified MnM2DNEO and modified MxM2DNEO with and without SMOTE, and after vascular calcification removal.

		Without SMOTE		With SMOTE		With SMOTE and vascular calcification removal	
Techniques	Database	Mean sensitivity (%)	FP per image	Mean sensitivity (%)	FP per image	Mean sensitivity (%)	FP per image
MxM2DNEO	DDSM	99	3.35	100	3.05	100	2.9
	INbreast	97.14	0.67	100	2.01	100	1.93
	PGIMER-IITKGP	100	2	100	1.77	100	1.61
MnM2DNEO	DDSM	100	2.59	100	2.55	100	2.34
	INbreast	100	1.78	100	1.61	100	1.57
	PGIMER-IITKGP	100	0.68	100	0.66	100	0.61

Table 4. Comparison of performances with respect to mean maximum sensitivity achieved by proposed and competing techniques and their corresponding mean FP per image with standard deviation.

	DDSM		INbreast		PGIMER-IITKGP	
Techniques, year	Mean sensitivity (%)	FP per image	Mean sensitivity (%)	FP per image	Mean sensitivity (%)	FP per image
SVM-SEL [11], 2002	99	8.51 ± 2.68	100	4.25 ± 3.05	92.89	1.59 ± 1.09
Unsharp masking [21], 2017	92	3.79 ± 0.86	97.14	2.13 ± 2.2	94.43	0.91 ± 0.45
Top hat transform [16], 2013	73	8.37 ± 2.45	97.14	2.29 ± 1.79	97.29	3.14 ± 2.67
MnM2DNEO [24], 2019	100	2.59 ± 1.47	100	1.78 ± 1.2	100	0.68 ± 0.62
MxM2DNEO [24], 2019	99	3.35 ± 1.89	97.14	0.67 ± 0.31	100	2 ± 2.81
Modified MnM2DNEO	100	2.34 ± 0.83	100	1.57 ± 1.17	100	0.61 ± 0.47
Modified MxM2DNEO	100	2.9 ± 1.25	100	1.93 ± 2.07	100	1.61 ± 0.87

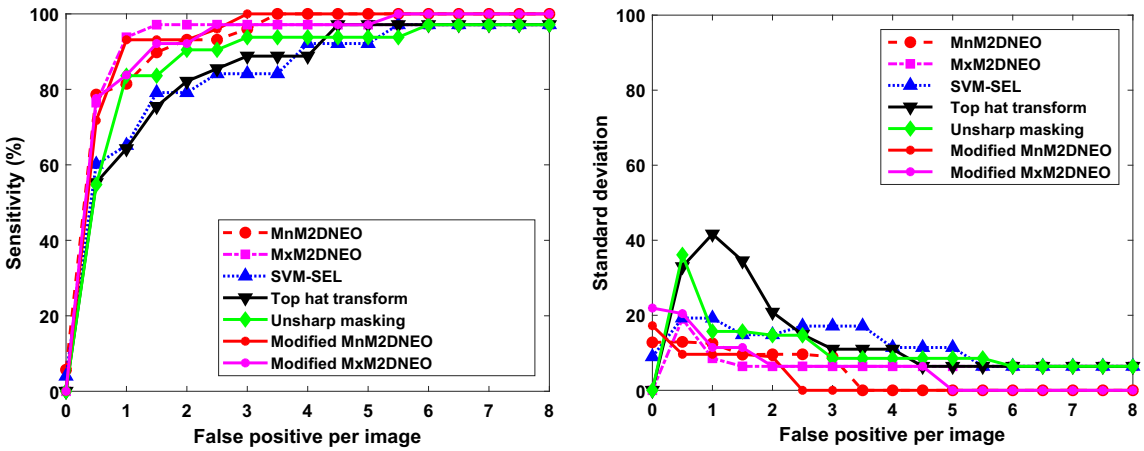


Figure 4. The comparison of performances of the proposed and competing techniques for the INbreast database using (a) FROC plots and (b) standard deviation of sensitivities at different operating points.

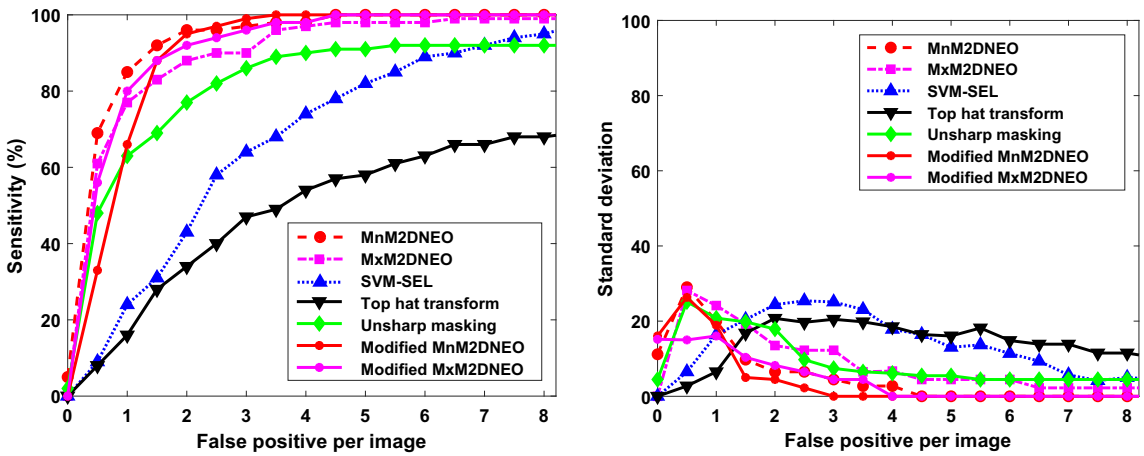


Figure 5. The comparison of performances of the proposed and competing techniques for the DDSM database using (a) FROC plots and (b) standard deviation of sensitivities at different operating points.

modified MxM2DNEO, as shown in figure 6. The deviation between the simulated results and the results reported for the competing techniques could be due to variation in the set of images and difference in evaluation criteria used in the study. We have used a more rational criterion to judge the true positive and FP clusters.

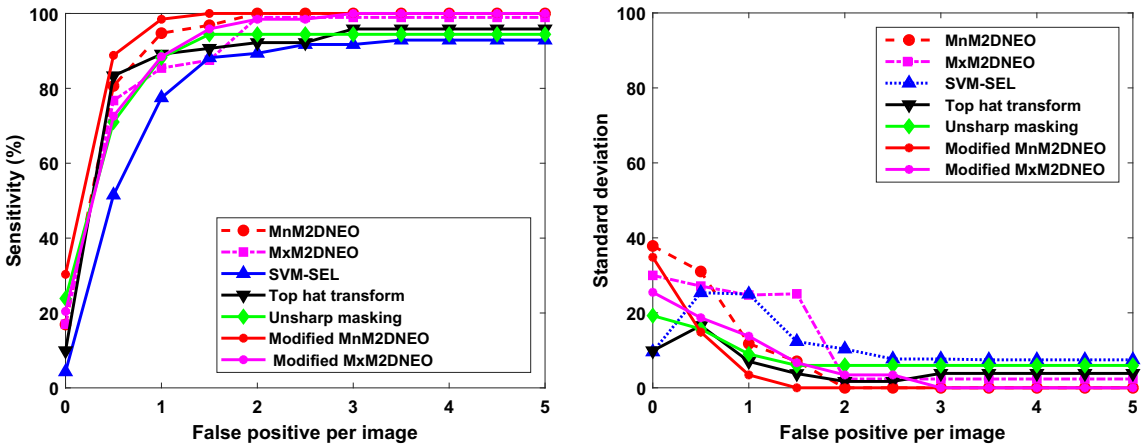


Figure 6. The comparison of performances of the proposed and competing techniques for the PGIMER-IITKGP database using (a) FROC plots and (b) standard deviation of sensitivities at different operating points.

Table 5. Comparison of performances of proposed and competing techniques with respect to sensitivity and the corresponding FP per image as reported in the literature.

Techniques, year	Database	Total test images	Sensitivity (%)	FP per image
Marrocco *et al* [40], 2010	Nijmegen	40	98	2.2
Gallardo-Caballero *et al* [41], 2012	DDSM	200	91.8	4.45
Liu *et al* [18], 2015	INbreast	410	92	2.3
Wang *et al* [22], 2017	Private	292	90	0.69
Wang and Yang [23], 2018	Private	292	87.5	0.5
Modified MxM2DNEO	DDSM	197	100	2.9
	INbreast	410	100	1.93
	PGIMER-IITKGP	110	100	1.61
Modified MnM2DNEO	DDSM	197	100	2.34
	INbreast	410	100	1.57
	PGIMER-IITKGP	110	100	0.61

The comparative results of proposed techniques with a few other competing techniques are presented in table 5. The competing techniques have reported their results using similar evaluation criteria as used to evaluate proposed techniques. As shown in table 5, proposed modified MnM2DNEO outperforms other competing techniques with respect to maximum sensitivity and corresponding FP per image. For the performance comparison of proposed MnM2DNEO, note that Marrocco *et al* achieved 98% sensitivity with 2.2 FP per image on a relatively smaller dataset containing 40 mammograms.

4. Conclusions

In this work, the techniques proposed by Karale *et al* [24], namely MnM2DNEO and MxM2DNEO, are improved by achieving 100% sensitivity at reduced FP rates. The performance of the classifier is improved by handling data imbalance effectively using SMOTE and under-sampling technique based on data distribution. This results in improvement of sensitivity and reduction of FPs. A simple vascular calcification removal technique is proposed for further reduction of FPs.

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel R L, Torre L A and Jemal A 2018 Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68: 394–424

[2] Nalawade Y V 2009 Evaluation of breast calcifications. *Indian J. Radiol. Imaging* 19: 282–286

[3] Pace L E and Keating N L 2014 A systematic assessment of benefits and risks to guide breast cancer screening decisions. *J. Am. Med. Assoc.* 311: 1327–1335

[4] Oeffinger K C, Fontham E T H, Etzioni R, Herzig A, Michaelson J S, Shih Y T, Walter L C, Church T R, Flowers C R, LaMonte S J, Wolf A M D, DeSantis C, Lortet-Tieulent J, Andrews K, Manassaram-Baptiste D, Saslow D, Smith R A, Brawley O W and Wender R 2015 Breast cancer screening for women at average risk: 2015 guideline update

from the American Cancer Society. *J. Am. Med. Assoc.* 314: 1599–1614

[5] Cox R F, Hernandez-Santana A, Ramdass S, McMahon G, Harmey J H and Morgan M P 2012 Microcalcifications in breast cancer: novel insights into the molecular mechanism and functional consequence of mammary mineralisation. *Br. J. Cancer* 106: 525–537

[6] Bulas D and Shah N 2014 International pediatric radiology education: who should be trained, and how? *Pediatr. Radiol.* 44: 639–641

[7] Papadopoulos A, Fotiadis D I and Costaridou L 2008 Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. *Comput. Biol. Med.* 38: 1045–1055

[8] Kim J K and Park H W 1999 Statistical textural features for detection of microcalcifications in digitized mammograms. *IEEE Trans. Med. Imaging* 18: 231–238

[9] Soltanian-Zadeh H, Rafiee-Rad F and Pourabdollah-Nejad D S 2004 Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms. *Pattern Recogn.* 37: 1973–1986

[10] Yu S and Guan L 2000 A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films. *IEEE Trans. Med. Imaging* 19: 115–126

[11] El-Naqa I, Yang Y, Wernick M N, Galatsanos N P and Nishikawa R M 2002 A support vector machine approach for detection of microcalcifications. *IEEE Trans. Med. Imaging* 21: 1552–1563

[12] Wei L, Yang Y, Nishikawa R M, Vernick M N and Edwards A 2005 Relevance vector machine for automatic detection of clustered microcalcifications. *IEEE Trans. Med. Imaging* 24: 1278–1285

[13] Peng R, Chen H and Varshney P K 2009 Noise-enhanced detection of micro-calcifications in digital mammograms. *IEEE J. Sel. Topics Signal Process.* 3: 62–73

[14] Nakayama R, Uchiyama Y, Yamamoto K, Watanabe R and Namba K 2006 Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms. *IEEE Trans. Biomed. Eng.* 53: 273–283

[15] Oliver A, Torrent A, Lladó X, Tortajada M, Tortajada L, Sentís M, Freixenet J and Zwiggelaar R 2012 Automatic microcalcification and cluster detection for digital and digitised mammograms. *Knowl. Based Syst.* 28: 68–75

[16] Zhang X, Homma N, Goto S, Kawasumi Y, Ishibashi T, Abe M, Sugita N and Yoshizawa M 2013 A hybrid image filtering method for computer-aided detection of microcalcification clusters in mammograms. *J. Med. Eng.*https://doi.org/10.1155/2013/615254

[17] Shin S, Lee S and Yun I D 2014 Classification based microcalcification detection using discriminative restricted boltzmann machine in digitized mammograms. In: *Proceedings of Medical Imaging*, SPIE, p. 90351L

[18] Liu X, Mei M, Liu J and Hu W 2015 Microcalcification detection in full-field digital mammograms with pfcm clustering and weighted SVM-based method. *EURASIP J. Adv. Signal Process.* 2015: 1

[19] Guo Y, Dong M, Yang Z, Gao X, Wang K, Luo C, Ma Y and Zhang J 2016 A new method of detecting micro-calcification clusters in mammograms using contourlet transform and non-linking simplified PCNN. *Comput. Methods Prog. Biomed.* 130: 31–45

[20] Mordang J, Janssen T, Bria A, Kooi T, Gubern-Mérida A and Karssemeijer N 2016 Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. In: *Proceedings of IWDM*, pp. 35–42

[21] Karale V A, Mukhopadhyay S, Singh T, Khandelwal N and Sadhu A 2017 Automated detection of microcalcification clusters in mammograms. In: *Proceedings of Medical Imaging*, SPIE, vol. 10134, p. 101342R

[22] Wang J, Nishikawa R M and Yang Y 2017 Global detection approach for clustered microcalcifications in mammograms using a deep learning network. *J. Med. Imaging* 4: 024501

[23] Wang J and Yang Y 2018 A context-sensitive deep learning approach for microcalcification detection in mammograms. *Pattern Recogn.* 78: 12–22

[24] Karale V A, Ebenezer J P, Chakraborty J, Singh T, Sadhu A, Khandelwal N and Mukhopadhyay S 2019 A screening CAD tool for the detection of microcalcification clusters in mammograms *J. Digit. Imaging* 32: 728–745

[25] Rose C, Turi D, Williams A, Wolstencroft K and Taylor C 2006 Web services for the DDSM and digital mammography research. In: *Proceedings of the 8th International Conference on Digital Mammography*, IWDM'06. Berlin–Heidelberg: Springer, pp. 376–383

[26] Moreira I C, Amaral I, Domingues I, Cardoso A, Cardoso M J and Cardoso J S 2012 INbreast: toward a full-field digital mammographic database. *Acad. Radiol.* 19: 236–248

[27] Guan P P and Yan H 2012 A hierarchical multilevel thresholding method for edge information extraction using fuzzy entropy. *Int. J. Mach. Learn. Cyb.* 3: 297–305

[28] Seth S and Mukhopadhyay S 2010 Multi-level thresholding-based breast segmentation in mammograms. In: *Proceedings of the International Conference on Communication, Computers, and Devices (ICCCD-2010)*, India

[29] Ciecholewski M 2017 Microcalcification segmentation from mammograms: a morphological approach. *J. Digit. Imaging* 30: 172–184

[30] Shen L, Rangayyan R M and Desautels J L 1992 Shape analysis of mammographic calcifications. In: *Proceedings of the Fifth Annual IEEE Symposium on Computer-Based Medical Systems*, IEEE, pp. 123–128

[31] Gonzalez R C and Woods R E 2002 *Digital image processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall

[32] Haralick R M, Shanmugam K and Dinstein I 1973 Textural features for image classification. *IEEE Trans. Syst. Man Cyb.* 3: 610–622

[33] Dalal N and Triggs B 2005 Histograms of oriented gradients for human detection. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, IEEE, vol. 1, pp. 886–893

[34] Ding C and Peng H 2005 Minimum redundancy feature selection from microarray gene expression data. *J. Bioinf. Comput. Biol.* 3: 185–205

[35] Velez D R, White B C, Motsinger A A, Bush W S, Ritchie M D, Williams S M and Moore J H 2007 A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. *Genet. Epidemiol.* 31: 306–315

[36] Chawla N V, Bowyer K W, Hall L O, and Kegelmeyer W P 2002 SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16: 321–357

[37] Cristianini N and Shawe-Taylor J 2000 *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press

[38] Woods K S, Solka J L, Priebe C E, Doss C C, Bowyer K W and Clarke L P 1993 Comparative evaluation of pattern recognition techniques for detection of microcalcifications. *Int. J. Pattern Recogn. Artif. Intell.* 7(6): 841–852

[39] Lai K C, Slanetz P J and Eisenberg R L 2012 Linear breast calcifications. *Am. J. Roentgenol.* 199: W151–W157

[40] Marrocco C, Molinara M, D'Elia C and Tortorella F 2010 A computer-aided detection system for clustered microcalcifications. *Artif. Intell. Med.* 50: 23–32

[41] Gallardo-Caballero R, García-Orellana C J, García-Manso A, González-Velasco H M and Macías-Macías M 2012 Independent component analysis to detect clustered microcalcification breast cancers. *Sci. World J.* 2012: 540457