Published in final edited form as:
Phys Med Biol. 2016 October 7; 61(19): 7092–7112. doi:10.1088/0031-9155/61/19/7092.

Analysis of computer-aided detection techniques and signal characteristics for clustered microcalcifications on digital mammography and digital breast tomosynthesis

Ravi K Samala, **Heang-Ping Chan**, **Lubomir M Hadjiiski**, and **Mark A Helvie**
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109-5842, rsamala@med.umich.edu

Abstract

With IRB approval, digital breast tomosynthesis (DBT) images of human subjects were collected using a GE GEN2 DBT prototype system. Corresponding digital mammograms (DMs) of the same subjects were collected retrospectively from patient files. The data set contained a total of 237 views of DBT and equal number of DM views from 120 human subjects, each included 163 views with microcalcification clusters (MCs) and 74 views without MCs. The data set was separated into training and independent test sets. The preprocessing, object prescreening and segmentation, false positive reduction and clustering strategies for MC detection by three computer-aided detection (CADe) systems designed for DM, DBT, and a planar projection image generated from DBT were analyzed. Receiver operating characteristic (ROC) curves based on features extracted from microcalcifications and free-response ROC (FROC) curves based on scores from MCs were used to quantify the performance of the systems. Jackknife FROC (JAFROC) and non-parametric analysis methods were used to determine the statistical difference between the FROC curves. The difference between the $\mathrm{CAD}_{\mathrm{DM}}$ and $\mathrm{CAD}_{\mathrm{DBT}}$ systems when the false positive rate was estimated from cases without MCs did not reach statistical significance. The study indicates that the large search space in DBT may not be a limiting factor for CADe to achieve similar performance as that observed in DM.

Keywords

digital breast tomosynthesis; digital mammography; computer-aided detection; microcalcification; planar projection image

I. INTRODUCTION

Mammography has been the standard method for breast cancer screening. Major technological innovations for mammography have been transitioning from screen-film mammography (SFM) to full-field digital mammography (DM) and currently to digital breast tomosynthesis (DBT). DM was shown to have similar overall performance as SFM (Lewin *et al.*, 2001; Lewin *et al.*, 2002; Skaane *et al.*, 2003; Skaane and Skjennald, 2004; Skaane *et al.*, 2005; Pisano *et al.*, 2005) with the advantages of lower recall rate, (Lewin *et al.*, 2001; Lewin *et al.*, 2002), better performance for women under the age of 50 years and women with heterogeneously dense or extremely dense breasts (Pisano *et al.*, 2005), and better image quality for microcalcifications (Fischer *et al.*, 2002; Kim *et al.*, 2006). DM provides the benefits of electronic storage and retrieval, image processing for signal enhancement, and contrast manipulation in real time. DBT generates a pseudo 3 dimensional (3D) volume of the breast, which is different from the typical 2D image presented by DM. Early prospective screening trials comparing DM with combined DM and DBT have demonstrated promising results. The Oslo study (Oslo Tomosynthesis Screening Trial: OTST) compared 12,631 examinations and concluded that the combined mode detected more invasive cancers and reduced pre-arbitration false positive (FP) scores (Skaane *et al.*, 2013a; Skaane *et al.*, 2013b). In the screening with tomosynthesis or standard mammography (STORM) study (Ciatto *et al.*, 2013), 7,292 examinations were analyzed. The study observed an increase in detection rate as well as a potential for reducing recalls by the combined DBT and DM reading. Several preliminary observer performance studies comparing DM and DBT for the same cases showed that the visibility of cancer in DBT is either equal or superior to DM (Andersson *et al.*, 2008), DBT has the potential to reduce recall rate (Gur *et al.*, 2009) and may offer greater benefits for less experienced radiologists (Wallis *et al.*, 2012). For non-calcified cancer, DBT was shown to have higher accuracy compared to DM (Ciatto *et al.*, 2013; Haas *et al.*, 2013; Rose *et al.*, 2013; Skaane *et al.*, 2013b). However, for evaluation of microcalcification clusters (MCs) there are mixed observations. Two studies showed that the microcalcifications in DBT had lower conspicuity (Poplack *et al.*, 2007) and detection sensitivity (Spangler *et al.*, 2011) than in DM. Another study showed that the detectability of microcalcifications in DBT was comparable to that in DM (Andersson *et al.*, 2008). A study by Kopans *et al* (Kopans *et al.*, 2011) found that the clarity of microcalcifications in DBT was comparable to DM in 50.4% of the cases while superior in 41.6% of the cases. These studies used relatively small data sets with different DM and/or different DBT systems so that the observations might be influenced by a number of factors and statistical variations. Further studies are needed to investigate the performance of DBT alone in comparison with DM in evaluation of MCs.

Computer-aided detection (CADe) of MCs for DM has been developed and used clinically for more than a decade (Lindfors *et al.*, 2006; Giger *et al.*, 2008). For DBT, CADe for MCs may potentially be even more useful because of the limitations in their visibility, which can be attributed to factors including large search space, lower signal-to-noise ratio due to cumulation of detector noise from multiple low-dose projections, blurring due to oblique incidence to the detector, potential patient motion due to relatively long scan time and reconstruction using thick slice spacing. For DBT systems with continuous x-ray source motion, focal spot blur may also affect the visibility of microcalcifications. DBT can reduce the tissue overlap and lead to better visibility than DM for soft tissue lesions such as masses, architectural distortion and focal asymmetry. It is also expected that future breast screening will entail usage of DBT alone in combination with ‘synthesized’ mammogram generated from DBT so as to avoid the additional dose in the combined DM and DBT approach. In this study, we compare the characteristics of CADe for DM and DBT for corresponding cases to evaluate the performance difference.

A brief overview of CADe for MCs on projection views or the reconstructed slices or volumes of DBT is shown in Table 1. For overview of the CADe work on DM the reader is referred to review articles (Giger *et al.*, 2008; Elter and Horsch, 2009). In this study, along with DM and DBT we also compared CADe for MCs on planar projection image (PPJ) generated from DBT volume enhanced with multiscale bilateral filtering (Lu *et al.*, 2015). PPJ is a novel representation of DBT volume on a single plane introduced by our group specifically for MC detection. PPJ contains only the high frequency information, which is different from the C-view or ‘synthesized’ mammogram. Our laboratory has previously developed CADe systems for MCs in screen-film mammography and digital mammography (Chan *et al.*, 1987; Chan *et al.*, 1990; Chan *et al.*, 1995; Gurcan *et al.*, 2002; Sahiner *et al.*, 2006; Ge *et al.*, 2006). We have been investigating enhancement (Lu *et al.*, 2012b; Lu *et al.*, 2012a) and detection of MCs (Sahiner *et al.*, 2012; Samala *et al.*, 2012; Samala *et al.*, 2013; Samala *et al.*, 2014b; Wei *et al.*, 2014) and mass (Chan, 2015; Samala *et al.*, 2016) in digital breast tomosynthesis. In this work, we quantitatively assess the differences for different steps between detection of MCs in DM, DBT reconstructed volume and PPJ image (fig. 1).

DBT has the advantage of reducing tissue overlap and initial reader studies have shown that this feature can lead to improved sensitivity and specificity. In this study, we attempted to understand the advantages and challenges for detection of MCs in DM and DBT from a CADe perspective. The following sections are arranged as follows: description of the data set, DBT reconstruction and PPJ image generation methodology, followed by the description of the three CADe systems for DM, DBT and PPJ; the results section presents the quantitative analysis of signal characteristics and cluster detection performance and conclude with the discussion of the observed findings.

II. MATERIALS AND METHODS

II.A. Data set

With Institutional Review Board (IRB) approval, DM and DBT cases were collected in the Department of Radiology at the University of Michigan Health System. The DM images were obtained during the patients’ clinical care and retrospectively collected from the patient files for this study. General Electric (GE) FFDM systems were used for DM imaging. With written informed consent, the DBT scans were acquired from patients who were recommended for biopsy of a suspicious lesion (BI-RADS assessment of 4 or 5) using a GE GEN2 DBT prototype system. The DBT system uses a step-and-shoot design with a total tomographic angle of 60° and an angular interval of 3°. The x-ray source consists of an Rh anode and Rh filter. The projection views (PVs) are recorded by a CsI/a:Si flat panel detector with a pixel pitch of 0.1 mm × 0.1 mm. The DBT volume for a given view was reconstructed from the 21 projection views at 1 mm slice spacing using simultaneous algebraic reconstruction technique (SART) (Zhang *et al.*, 2006) with multiscale bilateral filtering regularization (MSBF) (Lu *et al.*, 2015). A PPJ image was generated (see Section II.B) from each DBT volume (Samala *et al.*, 2014a). The PPJ image therefore has a one-to-one correspondence with the DBT volume.

Each case consisted of craniocaudal (CC) and mediolateral oblique (MLO) views of the DBT, PPJ, and the available corresponding DM of a breast. A total of 231 views with MCs and 76 views free of MCs were collected for this study. A DBT scan for one of the cases was lost due to technical failure of the system. Fig. 2 shows the views for each of the training and testing sets of the CADe systems. For all views with MCs, a Mammography Quality

Standards Act (MQSA) radiologist marked the cluster based on biopsy location information in the DM and DBT using a 2D and 3D bounding box, respectively. The radiologist also provided a subtlety rating of each marked cluster on a scale of 1–10 with 1 being most visible, and the BI-RADS breast density category of the breast. Fig. 3 shows the histograms of the subtlety ratings of the MCs and the breast density of the data set. The data set was partitioned by case into training and test sets, with the two views of the same case always assigned to the same set.

Three systems, $\mathrm{CAD}_{\mathrm{DBT}}$, $\mathrm{CAD}_{\mathrm{PPJ}}$ and $\mathrm{CAD}_{\mathrm{DM}}$ were used for the detection of MCs in DBT, PPJ, and DM, respectively. The $\mathrm{CAD}_{\mathrm{DBT}}$ and $\mathrm{CAD}_{\mathrm{PPJ}}$ systems were trained on 127 views and tested on 180 views. The training set of 127 views has 63 corresponding DM views and the test set of 180 views has 174 corresponding DM views. The $\mathrm{CAD}_{\mathrm{DM}}$ was previously trained and validated with a data set of 192 views (Ge *et al.*, 2006) that were acquired with a GE Senographe 2000D system and were independent of the cases collected in the current study. The $\mathrm{CAD}_{\mathrm{DM}}$ was not retrained using the current training set. For the discussion below, the training set refers to the 63 views in the current data set and the test set of 100 views with MCs and 74 views without MCs. Of the 237 DM views, 32 were imaged using a GE Senographe 2000D and 205 using GE Senographe Essential system. The term ‘view’ indicates the CC or MLO view image (DM or PPJ) or a DBT volume. A medical physicist experienced in mammography marked individual microcalcifications in 40 corresponding DM and DBT views from the training set to analyze the differences in the contrast-to-noise ratios (CNRs) of microcalcifications. We did not attempt to identify corresponding individual microcalcifications in DM and DBT so that the analysis was only performed statistically as a group.

II. B. DBT reconstruction and planar projection image generation

An MSBF regularized SART (Lu *et al.*, 2012a; Lu *et al.*, 2015) was developed for DBT reconstruction. Our study showed that MSBF regularization is superior to total variation regularization in that it enhances microcalcifications while preserving the appearance of mass spiculations and fibrous structures. At the end of each SART iteration, a reconstructed DBT slice is separated into high and low frequency bands using Laplacian pyramid decomposition. The high frequency bands containing microcalcifications and noise are selectively enhanced using bilateral filtering. The domain and range filter parameters of the bilateral filter are optimized to suppress noise while increasing the CNR of microcalcifications. The reverse process recombines all levels of the Laplacian pyramid into an enhanced DBT slice and used as input to the next iteration. Two iterations were used to reconstruct the DBT and generate the PPJ in this study.

For each DBT volume, the PPJ image is generated from the high-frequency component of the DBT slices at the final iteration. After MSBF is applied, maximum intensity projection is applied to the stack of high frequency components of the DBT slices perpendicular to the detector plane to produce the PPJ image for the view (Samala *et al.*, 2014a).

II. C. CADe systems

The overall strategy for detection of MCs in DBT, PPJ, or DM is to enhance the microcalcifications from the background, detect and segment microcalcification candidates, reduce FPs, and form clusters. The major similarities and differences between the CADe methods are shown in fig. 4. The details of the CADe systems have been described previously (Ge *et al.*, 2006; Samala *et al.*, 2014b; Samala *et al.*, 2014a). Each system is summarized briefly below.

II. C. 1. CADe for digital breast tomosynthesis and planar projection image— For $\mathrm{CAD_{DBT}}$, the DBT volume is first preprocessed using contrast-to-noise ratio (CNR) enhancement filters and multiscale calcification response (MCR). The MCR is the maximum response value of a response function calculated at multiple scales of the DBT volumes filtered with 3D Gaussian filters of a range of sizes. The response function is derived from the eigenvalues of the Hessian matrices and designed to have high values for objects that are approximately spherical in structure. The response values combined at multiple scales result in an MCR that enhances microcalcifications of various sizes. The CNR enhancement filter is a combination of three 2D boxcar filters designed to selectively enhance signals and suppress the background. The filter is convolved with the DBT volume slice-by-slice. The output of the MCR and CNR is combined voxelwise using a weighting function to generate an enhancement-modulated calcification response (EMCR) volume. The EMCR volume is subjected to a combination of iterative thresholding and region growing to detect the microcalcification candidates. In this process, the gray-level threshold automatically starts from the maximum value of the image histogram and segments objects above the threshold using 26-connectivity. The threshold is iteratively reduced until a predetermined number of MC candidates are detected. These objects serve as seeds for a refined segmentation, in which the local noise level and CNR of each voxel in a region around the seed is calculated adaptively and all connected voxels by 26-connectivity and greater than a CNR threshold value are segmented to form a microcalcification candidate. The maximum CNR value (CNR_{obj}) and the size of the segmented object are determined and a set of adaptive thresholds of CNR_{obj} are estimated from histogram analysis of the candidate objects in the DBT volume. During the cluster detection stage, the top N objects with the highest CNR_{obj} are used as cluster seeds and the rest as potential cluster members. Dynamic conditional clustering starts from the cluster seed list and iteratively grows a cluster. After all potential clusters are formed, a decision-tree classifier that utilizes the adaptive thresholds of CNR_{obj}, the size of the individual member objects, the number of objects in the cluster is employed to reduce FP clusters. As a final step, the cluster shape and object shape obtained by maximum intensity projection within the cluster are used to exclude FPs due to vascular or fibrous artifacts.

For $\mathrm{CAD_{PPJ}}$, since the low frequency bands in the Laplacian pyramid are not used during the generation of the PPJ image, no preprocessing is needed for removing the background. The iterative thresholding and region growing for identifying candidates and the local object segmentation for refinement of the object, similar to those described above for the DBT volume except that 8-connectivity is used for a 2D image, are directly applied to the PPJ image to identify the microcalcification candidates. In a PPJ image, the microcalcification candidates are screened for FPs using a convolution neural network (CNN) trained to differentiate microcalcifications from linear tissue structures and artifacts before dynamic conditional clustering. FP reduction is performed with a similar approach to that in CAD$_{\mathrm{DBT}}$ using a decision tree with adaptive CNR_{obj} thresholds. All parameters were specifically designed for PPJ images.

II.C.2. CADe for Digital Mammography—For CAD$_{\mathrm{DM}}$, the raw (“For processing”) DM image is preprocessed using an inverted logarithmic transformation and microcalcifications are enhanced using the CNR enhancement filter. The microcalcification candidates are identified by the combined iterative thresholding and region growing process, as described above. Refined segmentation of the candidate objects is performed by calculating the CNR values adaptively in the local region similar to the process described in Section II.C.1. The objects are screened using rule-based and CNN classifiers. Clusters are formed by dynamic regional clustering in a hierarchical order based on the number of objects from high to low in a local region. Morphological and CNN features extracted from the clusters are combined by a linear discriminant analysis (LDA) classifier into a decision variable to differentiate true and false MCs.

II. E. Evaluation methods

II. E. 1. Object-level performance evaluation—Signal characteristics are analyzed at microcalcification object level. For object-level detection, i.e., the detection of individual microcalcification candidates, the 2D image or 3D volume is enhanced to separate the microcalcification candidates from the background breast tissue. Once the candidates are detected, the candidate features can be used to assess the performance of enhancement and the ability to reduce the FPs. We observed that a high number of FPs at the candidate detection stage usually result in a high FP rate at the cluster detection stage. For the CAD$_{\mathrm{DM}}$ system, the most important object-level features are the CNN and CNR values of the detected objects. These features are useful for segmentation, FP reduction of individual microcalcification candidates and as predictor variables in the LDA function to generate a discriminant score. For CAD$_{\mathrm{DBT}}$ and CAD$_{\mathrm{PPJ}}$ systems, the CNR score of an object (CNR_{obj}) is the most important feature for segmentation, classification of individual microcalcifications and classification of clusters. For all three CADe systems, the area under the receiver operating characteristic (ROC) curve (AUC) value is used for evaluation of the performance of object-level TP and FP classification.

II. E. 2. Cluster-level performance evaluation—After the microcalcification candidates are grouped to form MCs, cluster-based FP reduction methods are used to further reduce FPs. The performance of a CADe system is analyzed using a free-response ROC (FROC) curve. The FROC curve is used to assess the performance of a CADe system where the average FP rate is estimated from cases either with or without MCs. For CAD$_{\mathrm{DBT}}$ and CAD$_{\mathrm{PPJ}}$, the decision variable is the maximum of the CNR_{obj} of all objects within a cluster. For CAD$_{\mathrm{DM}}$, the output of the LDA classifier is used as a decision variable. To assess the performance difference between the CADe systems using the FROC data, two approaches are used: (a) Jackknife free-response operating characteristic (JAFROC) (Chakraborty, 2008) where the figure-of-merit is the area under the alternative FROC curve with the FP rate estimated from cases without MCs. JAFROC estimates the probability of a marked cluster having a score higher than the cluster on the normal views. (b) Non-parametric analysis (Samuelson *et al.*, 2007; Samuelson and Petrick, 2006) where the figure-of-merit is the area under the FROC curve with the FP rate estimated from cases with MCs. This method uses permutation test to analyze the correlated FROC data at a threshold of number of FPs per view.

III. RESULTS

III. A. Comparison of object-level features

The number of microcalcification candidates extracted from the combination of iterative thresholding and region growing was chosen to be a maximum of 180, 800 and 300 for DM, DBT and PPJ, respectively. These numbers were experimentally designed for DM and based on the rank-sensitivity plot (Sahiner *et al.*, 2012; Samala *et al.*, 2014b) for DBT and PPJ. The evaluation at this stage is to assess the ability of the pre-processing methods in separating the candidates from the background and ranking them. A rank-sensitivity plot is similar to an FROC curve but is used to evaluate detection performance at the prescreening stage. The x-axis of the rank-sensitivity plot is the rank threshold R of the highest rank objects based on the EMCR score that are kept as cluster seeds at prescreening. The y-axis is the fraction (F) of true clusters in the data set that have at least one cluster seed within the bounding box of the true cluster at the rank threshold R. Comparing these plots for different parameter settings, the number of objects to be obtained to achieve the highest sensitivity at the prescreening stage is estimated.

The object-level features were compared using a subset of 40 training views (fig. 2, from the 63 training views). In this subset, 545 and 324 microcalcifications are manually marked in the DM images and DBT volumes, respectively. The CADe-detected objects that did not overlap with the marked locations are considered FPs. The locations marked in the DBT volume were projected onto the PPJ image to obtain the same number of marked locations for PPJ. The CNN and CNR features of the $\mathrm{CAD}_{\mathrm{DM}}$ and the CNR features of $\mathrm{CAD}_{\mathrm{DBT}}$ and $\mathrm{CAD}_{\mathrm{PPJ}}$ are compared using ROC curves shown in fig. 5. The AUC of the CNR features is 0.84±0.01, 0.85±0.01 and 0.90±0.007 for DM, DBT and PPJ respectively. The AUC values of the CNN feature for DM is 0.86±0.01. The ROC curves were fitted with a “proper” binormal model using ROC analysis program (ROC library 1.0.3 v 2011, URL: http://xray.bsd.uchicago.edu/krl/) by Metz *et al* (Metz and Pan, 1999). Note that the training set with manual microcalcification markings is used for only object-level feature analysis, the rest of the analysis and performance comparison were conducted on the test set.

III. B. Comparison of cluster-level performance

The number of microcalcification candidates in a cluster is an important feature that has been consistently used for CADe of MCs. The performance of this feature is compared using the ROC curves of the test set shown in fig. 6.

For evaluation of cluster detection performance, if the centroid of a detected cluster is within the radiologist-marked bounding box of a true cluster, the cluster is marked as TP. The view-based FROC curves of the three CADe systems for the test set are shown in Fig. 7. In view-based evaluation, the MC on each view is considered an independent target and the sensitivity is determined as the fraction of true MCs in the test set that are detected. The results from statistical significance testing are shown in Tables II and III. At 85% sensitivity, the FP rates of the CAD_{DM}, CAD_{DBT}, and CAD_{PPJ} systems were 5.74, 2.19, 1.71 FPs/view, respectively, when the average FP rate was estimated from views with MCs. The result of significance testing by the non-parametric analysis method (Table II) shows that the differences between the performances of CAD_{DM} versus CAD_{DBT} and CAD_{PPJ} were statistically significant at an operating threshold of 1 FP/view. When the average FP rate was estimated from the views without MCs, which is the majority of the screening population, at 85% sensitivity, the FP rates of the CAD_{DM}, CAD_{DBT}, and CAD_{PPJ} systems were 6.10, 2.19 and 0.93 FPs/view, respectively. JAFROC analysis in Table III shows that the difference between CAD_{DBT} and CAD_{PPJ} was statistically significant with a *p*-value of 0.034. The *p*-value for the difference between DM and PPJ just fell short of significance with a *p*-value of 0.059. The difference between CAD_{DM} and CAD_{DBT} did not achieve statistical significance.

We evaluated the effect of breast density on the detection performances of the CAD_{DBT}, CAD_{PPJ}, and CAD_{DM} systems. Because of the small number of cases when they were separated into four BI-RADS categories, we grouped the BI-RADS 1, 2 cases into a low density category and the BI-RADS 3, 4 cases into a high density category based on the experienced radiologist's density ratings on DMs. For the test set of 174 views, 28 views with MCs and 42 views without MCs were classified into the BI-RADS 1, 2 group, while 72 views with MCs and 32 views without MCs were classified into the BI-RADS 3, 4 group. The same grouping of cases was applied to both the DM and DBT images. Fig. 8 shows the performance differences between the two breast density groups for the three CADe systems.

IV. DISCUSSION

DBT has been integrated into clinical use in recent years. It is essential to investigate the advantages or limitations of DBT when compared to DM. A number of preliminary studies have shown that DBT has the potential to detect more invasive cancers. Since CADe was routinely used for screening mammography, it can be expected that CADe may be useful for DBT interpretation, especially for MC detection. We analyzed the differences of MC detection among DM, DBT and PPJ using corresponding DM and DBT from the same patients.

The overall structure of a typical CADe system for MC detection includes (i) preprocessing, (ii) microcalcification candidate detection, (iii) candidate-based FP reduction, (iv) clustering, and (v) cluster-based FP reduction as shown in fig. 4. Although the general structure is similar between the CADe systems, each system uses techniques specifically trained according to the properties of each type of images. Due to the large search space in DBT, the preprocessing step in CAD_{DBT} is more complex than that in CAD_{DM}. Contrast- and shape-based enhancement of candidates in the DBT volume and weighted fusion of the enhanced volumes are used to effectively improve the signal strength of microcalcifications. For DM, the CNR enhancement filter is used to enhance candidates and remove the low-frequency tissue background on the DM image. Since a PPJ image is derived from the high frequency band of the decomposed bilateral filtering regularized DBT volume, the low frequency background is removed during the PPJ generation process, essentially eliminating the need for preprocessing steps. All three CADe systems use a combination of iterative thresholding and region growing followed by adaptive local gray level thresholding for segmentation. For the CAD_{DBT} and CAD_{PPJ} systems, the number of potential candidates to be obtained in the prescreening stage is estimated experimentally based on the rank-sensitivity plots from the training set. Because of the larger search space, the number of candidates for DBT is much higher, resulting in more complex FP reduction methods in the subsequent steps. Object-based analysis is performed at this stage to evaluate the sensitivity of detecting microcalcifications while maintaining a low number of FPs. Fig. 5 shows the ability to discriminate between TP and FP objects in a training set of 40 views. For the CNR feature, CAD_{DBT} and CAD_{DM} are comparable and CAD_{PPJ} has the highest AUC showing a significant advantage to identify true microcalcifications. The CNR and CNN features of objects in DM are comparable in their discriminatory ability for distinguishing true and false microcalcifications. CAD_{DM} uses rule-based classifiers based on size and CNR of the objects detected to remove small objects, artifacts and large benign microcalcifications. For both CAD_{DM} and CAD_{PPJ}, a CNN was trained to reduce FP objects. In case of CAD_{DM}, the CNN was trained to separate microcalcifications from noisy linear structures. For CAD_{PPJ}, the CNN was trained to separate microcalcifications from linear structures, metallic-clip and blurring artifacts. No CNN was trained for FP reduction methods for DBT because of the contrast-and-shape-based enhancement at the preprocessing step.

To form clusters from the individual objects, dynamic conditional clustering is used in the CAD_{DBT} and CAD_{PPJ} systems to grow a cluster from the cluster seeds, and the priority is determined by the CNR_{obj} ranking of the seed. CNR thresholds in CAD_{DBT} and CAD_{PPJ} are generated based on the histogram of the CNR_{obj} of all the prescreened candidates and the CNR, size, and the number of the objects in the cluster are used to differentiate true and false clusters in a decision tree classifier. CAD_{DBT} has additional FP reduction steps using the shape of a cluster and a maximum-intensity-projection-based method. The highest CNR_{obj} of all the candidates within an MC is used as decision variable to generate the FROC curves for CAD_{DBT} and CAD_{PPJ}. CAD_{DM} uses regional clustering in which the priority of forming cluster is determined by the number of the candidates within a local region on the image. For cluster-based FP reduction, CAD_{DM} uses a trained LDA classifier to reduce FPs and the discriminant score is used as decision variable to generate the FROC curve. From the FROC curves for the test set where the average FP rate is estimated from views without MCs (fig. 7(b)), in the region below about 65% sensitivity where the three curves intersect, DBT and PPJ have similar performance but inferior to DM. This indicates that the features selected by the LDA classifier in DM are more effective in differentiating obvious MCs from FPs. The LDA uses the number of calcifications in a cluster as one of the features (fig. 6). This feature has substantially higher discriminating power for DM than for DBT and PPJ. After the crossover point, the CAD_{PPJ} performed better than the other two, having 85% sensitivity at 0.93 FPs/view. JAFROC in Table III shows that CAD_{PPJ} is significantly better than CAD_{DBT} with a p-value of less than 0.05. The difference between CAD_{PPJ} and CAD_{DM} approaches significance with a p-value of 0.059. From the FROC curves for the test set with average FP rate calculated from views with MCs (fig. 7(a)),

CAD_{DM} shows a substantial difference from the other systems up to approximately 65% sensitivity. The non-parametric analysis in Table II shows that the difference between CAD_{DM} and the other CAD systems was statistically significant when the operating threshold was set to 1 FP/view but CAD_{DM} reached a maximum of only 81% sensitivity at higher FP rates, which was lower than the maximum sensitivities by CAD_{DBT} and CAD_{PPJ} systems. No significant difference was observed between CAD_{DBT} and CAD_{PPJ}.

Analysis of the reasons why some of the MCs were missed in some systems and some were missed in all three systems studied is difficult because of the complexity of the CADe methods and the relatively small data set. MCs could be missed due to fewer number of candidates detected, over or under-segmentation, failure to recognize TPs by CNN, the CNR_{obj} falling below the thresholds, or loss of clusters due to cluster-based FP reduction steps. To gain an understanding of the performance of the CADe systems, we compared their performances in terms of the MC subtlety ratings by the experienced breast radiologist. The subtlety rating ranges from 1 to 10, with 1 being most visible. From the test set, CAD_{DM} lost 16 MCs, CAD_{DBT} lost 9 and CAD_{PPJ} lost 12, as shown in the analysis in figure 9. Four of the lost MCs were missed by all systems. Most of the missed clusters are subtle with scores higher than 4 indicating that these are generally difficult cases to detect.

Breast density has been shown to affect the performance of CADe systems for detection of mass and MCs on DM (Brem *et al.*, 2005). Similar results were observed in this study as shown in fig. 8. All the systems showed better MC detection performance in BI-RADS 1 and 2 breasts. Fig. 10 shows an example of a ductal carcinoma in situ (DCIS) for which the MC had a subtlety rating of 5 in DBT and the cluster was relatively conspicuous in both the DBT and the PPJ images. The same MC on the DM image had a subtlety rating of 9 due to the dense overlapping tissue with a BI-RADS density rating of 3. The DCIS was detected by CAD_{DBT} and CAD_{PPJ} but missed by CAD_{DM}. Fig. 11 shows a benign cluster missed by all the CADe systems. The case has a BI-RADS density rating of 3 in DBT. Further study with a larger data set will be needed to evaluate the extent to which breast density will affect MC detection in DM, DBT and PPJ.

The JAFROC analysis shows that the difference between the comparison of the pair of CAD_{DM}-CAD_{DBT} and CAD_{DM}-CAD_{PPJ} systems did not achieve statistical significance. However, the individual FROC curves indicated that CAD_{DBT} and CAD_{PPJ} could reach higher sensitivity at larger FP rates compared to CAD_{DM}. For CAD_{DBT}, a decision-tree like approach was designed to stratify the cluster candidates into different classes. This approach allows the flexibility of using different criteria for FP reduction in each class, rather than applying a decision threshold to a single discriminant score for all clusters (Samala *et al.*, 2014a; Samala *et al.*, 2014b). This may have reduced the false negatives for the subtle clusters.

The computation time for MC detection by CAD_{DM} and CAD_{PPJ} takes an average of less than 30 seconds per view while that by CAD_{DBT} takes an average of 3.8 minutes per view on a workstation with a CPU of 3.47 GHz and 24 GB of system memory. It may be noted that these CAD systems were implemented for experimental use and no attempt has yet been made to optimize the computation efficiency. In addition, the computation time does not influence radiologist's reading time because CAD processing generally is performed before the case is displayed and read by radiologists rather than concurrent with radiologist's reading.

This study has a number of limitations. First, the DM and DBT were not acquired under the same compression and with the same imaging system. We did not request additional DM from the subject at the time of DBT imaging to minimize radiation dose, and clinical combo (DBT plus DM) cases were not available to us for this study. The variations in the compression and the difference in the imaging techniques may affect the detection of MCs to some extent. However, the variations between the DM and DBT images could go in either directions and might not specifically bias one modality or the other when the results were averaged over the cases. Second, the CAD_{DBT} and CAD_{PPJ} systems were trained with the training set described in the current study (fig. 3(b)). The CAD_{DM} system was trained previously with a different data set that did not have corresponding DBT (Ge *et al.*, 2007). We did not retrain the CAD_{DM} system because the available DM in the current training set was smaller than the previous training set (Fig. 2). It is difficult to assess the effect of different training sets but in general more robust training should be achieved with a larger training set (Fukunaga and Hayes, 1989; Chan *et al.*, 1999). Third, the data set in the current study is limited in size. In particular, it is not large enough to assess the differences between malignant and benign MCs. Nevertheless, all MCs in the data set had undergone diagnostic workup and were recommended for biopsy so that they were suspicious enough to warrant detection during screening.

V. CONCLUSION

DBT has the potential to replace DM for breast cancer screening in the near future. For MC detection, previous studies indicate that it is challenging and time consuming for radiologists to visually search for subtle MC clusters in the large volume of DBT. In this study, we compared the detection of MCs in DM and DBT or PPJ using a data set of corresponding cases. The results show that computerized detection of MC in DBT or PPJ is comparable to that in DM, indicating computer vision may not be subject to the same problems of visual search by human eyes in large image space for subtle MCs. Although in the current study the DM and DBT were not acquired under the same compression, the final analysis was provided based on results from the entire data set and not based on cluster-by-cluster comparison. In other words, the comparison was made based on the mean response of the system across different types of MCs from different breast density categories and over the variations in image quality due to different compression in individual cases. Further studies will be conducted to improve the performance of CAD_{DBT} and CAD_{PPJ}, and evaluate whether CADe can assist radiologists in MC detection in DBT.

Acknowledgments

This work is supported by National Institutes of Health grant RO1 CA151443. We would like to thank Frank W. Samuelson and Nicholas Petrick for providing the code and assistance with the non-parametric analysis method.

REFERENCES

Andersson I, Ikeda DM, Zackrisson S, Ruschin M, Svahn T, Timberg P, Tingberg A. Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and BIRADS classification in a population of cancers with subtle mammographic findings. European Radiology. 2008; 18:2817–2825. [PubMed: 18641998]

Bernard S, Muller S, Onativia J. Computer-aided microcalcification detection on digital breast tomosynthesis data: A preliminary evaluation. Lect. Notes Comput. Sci. 2008; 5116:151–157.

Brem RF, Hoffmeister JW, Rapelyea JA, Zisman G, Mohtashemi K, Jindal G, DiSimio MP, Rogers SK. Impact of breast density on computer-aided detection for breast cancer. American Journal of Roentgenology. 2005; 184:439–444. [PubMed: 15671360]

Chakraborty DP. Validation and Statistical Power Comparison of Methods for Analyzing Free-response Observer Performance Studies. Academic Radiology. 2008; 15:1554–1566. [PubMed: 19000872]

Chan, H-P. Chapter 4. Detection and Diagnosis of Breast Mass in Digital Tomosynthesis: Computer-Aided Detection and Diagnosis in Medical Imaging. Li, Q.; Nishikawa, R., editors. Taylor & Francis Group, LLC: CRC Press, Boca Raton, FL; 2015.

Chan H-P, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM. Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography. Medical Physics. 1987; 14:538–548. [PubMed: 3626993]

Chan H-P, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam KL, Ogura T, Wu Y, MacMahon H. Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis. Investigative Radiology. 1990; 25:1102–1110. [PubMed: 2079409]

Chan H-P, Lo SCB, Sahiner B, Lam KL, Helvie MA. Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network. Medical Physics. 1995; 22:1555–1567. [PubMed: 8551980]

Chan H-P, Sahiner B, Wagner RF, Petrick N. Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers. Medical Physics. 1999; 26:2654–2668. [PubMed: 10619251]

Chan H-P, Sahiner B, Wei J, Hadjiiski LM, Zhou C, Helvie MA. Digital breast tomosynthesis: Computerized detection of microcalcifications in reconstructed breast volume using a 3D approach. Proc. SPIE. 2010; 7624:1D–D8.

Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, Tuttobene P, Bricolo P, Fantò C, Valentini M, Montemezzi S, Macaskill P. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. The Lancet Oncology. 2013; 14:583–589. [PubMed: 23623721]

Elter M, Horsch A. CADx of mammographic masses and clustered microcalcifications: A review. Medical Physics. 2009; 36:2052–2068. [PubMed: 19610294]

Fischer U, Baum F, Obenauer S, Luftner-Nagel S, Von Heyden D, Vosshenrich R, Grabbe E. Comparative study in patients with microcalcifications: full-field digital mammography vs screen-film mammography. European radiology. 2002; 12:2679–2683. [PubMed: 12386757]

Fukunaga K, Hayes RR. Effects of sample size on classifier design. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989; 11:873–885.

Ge J, Hadjiiski LM, Sahiner B, Wei J, Helvie MA, Zhou C, Chan H-P. Computer-aided detection system for clustered microcalcifications: comparison of performance on full-field digital mammograms and digitized screen-film mammograms. Physics in Medicine and Biology. 2007; 52:981–1000. [PubMed: 17264365]

Ge J, Sahiner B, Hadjiiski LM, Chan H-P, Wei J, Helvie MA, Zhou C. Computer aided detection of clusters of microcalcifications on full field digital mammograms. Medical Physics. 2006; 33:2975–2988. [PubMed: 16964876]

Giger ML, Chan H-P, Boone J. Anniversary Paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM. Medical Physics. 2008; 35:5799–5820. [PubMed: 19175137]

Gur D, Abrams GS, Chough DM, Ganott MA, Hakim CM, Perrin RL, Rathfon GY, Sumkin JH, Zuley ML, Bandos AI. Digital Breast Tomosynthesis: Observer Performance Study. American Journal of Roentgenology. 2009; 193:586–591. [PubMed: 19620460]

Gurcan MN, Chan H-P, Sahiner B, Hadjiiski L, Petrick N, Helvie MA. Optimal neural network architecture selection: Improvement in computerized detection of microcalcifications. Academic Radiology. 2002; 9:420–429. [PubMed: 11942656]

Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of Tomosynthesis Plus Digital Mammography and Digital Mammography Alone for Breast Cancer Screening. Radiology. 2013; 269:694–700. [PubMed: 23901124]

Kim HH, Pisano ED, Cole EB, Jiroutek MR, Muller KE, Zheng Y, Kuzmiak CM, Koomen MA. Comparison of calcification specificity in digital mammography using soft-copy display versus screen-film mammography. American Journal of Roentgenology. 2006; 187:47–50. [PubMed: 16794154]

Kopans D, Gavenonis S, Halpern E, Moore R. Calcifications in the Breast and Digital Breast Tomosynthesis. Breast Journal. 2011; 17:638–644. [PubMed: 21906207]

Lewin JM, D'Orsi CJ, Hendrick RE, Moss LJ, Isaacs PK, Karellas A, Cutter GR. Clinical comparison of full-field digital mammography and scree-film mammography for detection of breast cancer. AJR. 2002; 179:671–677. [PubMed: 12185042]

Lewin JM, Hendrick RE, D'Orsl CJ, Isaacs PK, Moss LJ, Karellas A, Slsney GA, Kuni CC, Cutter GR. Comparison of full-field digital mammography with screen-film mammography for cancer detection: results of 4,945 paired examinations. Radiology. 2001; 218:873–880. [PubMed: 11230669]

Lindfors KK, McGahan MC, Rosenquist CJ, Hurlock GS. Computer-aided detection of breast cancer: A cost-effectiveness study. Radiology. 2006; 239:710–717. [PubMed: 16569787]

Lu Y, Chan H-P, Samala RK, Hadjiiski L, L CP. Multiscale bilateral regularization in digital breast tomosynthesis (DBT). RSNA Program Book. 2012a:SSE22.

Lu, Y.; Chan, H-P.; Wei, J.; Hadjiiski, LM. Improving Image Quality of Digital Breast Tomosynthesis by Artifact Reduction; Proc. 11th Int. Workshop on Digital Mammography (Lecture Notes in Computer Science vol 7361); 2012b. p. 745-752.

Lu Y, Chan H-P, Wei J, Hadjiiski LM, Samala RK. Multiscale Bilateral Filtering for Improving Image Quality in Digital Breast Tomosynthesis. Medical Physics. 2015; 42:182–195. [PubMed: 25563259]

Metz CE, Pan X. "Proper" binormal ROC curves: Theory and maximum-likelihood estimation. Journal of Mathematical Psychology. 1999; 43:1–33. [PubMed: 10069933]

Park SC, Zheng B, Wang XH, Gur D. Applying a 2D based CAD scheme for detecting micro-calcification clusters using digital breast tomosynthesis images: An assessment. Proc. SPIE. 2008; 6915:691507.

Peters G, Muller S, Bernard S, Iordache R, Bloch I. Reconstruction-independent 3D CAD for mass detection in Digital Breast Tomosynthesis using fuzzy particles. Proc SPIE. 2006; 6144:Z1441.

Pisano ED, Gatsonis C, Hendrick E, Yaffe M. Diagnostic performance of digital versus film mammography for breast-cancer screening. The New England Journal of Medicine. 2005; 353:1773–1783. [PubMed: 16169887]

Poplack SP, Tosteson TD, Kogel CA, Nagy HM. Digital Breast Tomosynthesis: Initial Experience in 98 Women with Abnormal Digital Screening Mammography. American Journal of Roentgenology. 2007; 189:616–623. [PubMed: 17715109]

Reiser I, Nishikawa RM, Edwards AV, Kopans DB, Schmidt RA, Papaioannou J, Moore RH. Automated detection of microcalcification clusters for digital breast tomosynthesis using projection data only: A preliminary study. Medical Physics. 2008; 35:1486–1493. [PubMed: 18491543]

Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R. Implementation of Breast Tomosynthesis in a Routine Screening Practice: An Observational Study. American Journal of Roentgenology. 2013; 200:1401–1408. [PubMed: 23701081]

Sahiner B, Chan H-P, Hadjiiski LM, Helvie MA, Paramagul C, Ge J, Wei J, Zhou C. Joint two-view information for computerized detection of microcalcifications on mammograms. Medical Physics. 2006; 33:2574–2585. [PubMed: 16898462]

Sahiner B, Chan H-P, Hadjiiski LM, Helvie MA, Wei J, Zhou C, Lu Y. Computer-aided detection of clustered microcalcifications in digital breast tomosynthesis: A 3D approach. Medical Physics. 2012; 39:28–39. [PubMed: 22225272]

Samala RK, Chan H-P, Lu Y, Hadjiiski L, Wei J, Helvie MA. Computer-aided Detection of Clustered Microcalcifications in Digital Breast Tomosynthesis: False Positive Reduction Using Bilateral Filtering. RSNA Program Book. 2012:SSA19.

Samala RK, Chan H-P, Lu Y, Hadjiiski L, Wei J, Helvie MA. Digital breast tomosynthesis: computer-aided detection of clustered microcalcifications on planar projection images. Physics in medicine and biology. 2014a; 59:7457–7477. [PubMed: 25393654]

Samala RK, Chan H-P, Lu Y, Hadjiiski L, Wei J, Sahiner B, Helvie MA. Computer-aided detection of clustered microcalcifications in multiscale bilateral filtering regularized reconstructed digital breast tomosynthesis volume. Medical Physics. 2014b; 41:021901. [PubMed: 24506622]

Samala RK, Chan H-P, Lu Y, Hadjiiski LM, Wei J, Helvie MA. Computer-aided detection system for clustered microcalcifications in digital breast tomosynthesis using joint information from volumetric and planar projection images. Physics in medicine and biology. 2015; 60:8457. [PubMed: 26464355]

Samala RK, Chan H-P, Lu Y, Hadjiyski L, Wei J, Sahiner B, Helvie M. Detection of microcalcifications in breast tomosynthesis reconstructed with multiscale bilateral filtering regularization. Proc. SPIE. 2013; 8670:86701L–86701L-8.

Samala RK, Wei J, Chan H-P, Hadjiiski LM, Cha K, Helvie MA. First and second-order features for detection of masses in digital breast tomosynthesis. SPIE Medical Imaging. 2016; 9785:978523–978527.

Samuelson, FW.; Petrick, N. Comparing image detection algorithms using resampling; Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on; 2006. p. 1312-1315.

Samuelson, FW.; Petrick, N.; Paquerault, S. Advantages and examples of resampling for CAD evaluation; Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on; 2007. p. 492-495.

Skaane P, Balleyguier C, Diekmann F, Diekmann S, Piguet J-C, Kari Young and Niklason LT. Breast lesion detection and classification: Comparison of screen-film mammography and full-field digital mammography with soft-copy reading—Observer performance study. Radiology. 2005; 237:37–44. [PubMed: 16100086]

Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, Izadi M, Jebsen IN, Jahr G, Krager M. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology. 2013a; 267:47–56. [PubMed: 23297332]

Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, Izadi M, Jebsen IN, Jahr G, Krager M, Hofvind S. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. European Radiology. 2013b; 23:2061–2071. [PubMed: 23553585]

Skaane P, Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program -- The Oslo II Study. Radiology. 2004; 232:197–204. [PubMed: 15155893]

Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading --- Oslo I Study. Radiology. 2003; 229:877–884. [PubMed: 14576447]

Spangler ML, Zuley ML, Sumkin JH, Abrams G, Ganott MA, Hakim C, Perrin R, Chough DM, Shah R, Gur D. Detection and Classification of Calcifications on Digital Breast Tomosynthesis and 2D Digital Mammography: A Comparison. American Journal of Roentgenology. 2011; 196:320–324. [PubMed: 21257882] van Schie G, Karssemeijer N. Noise model for microcalcification detection in reconstructed tomosynthesis slices. Proc. SPIE. 2009; 7260:72600M.

Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: High-resolution x-ray imaging observer study. Radiology. 2012; 262:788–796. [PubMed: 22274840]

Wei J, Chan H-P, Hadjiiski LM, Helvie MA, Lu Y, Zhou C, Samala RK. Multi-channel response analysis on 2D projection views for detection of clustered microcalcifications in digital breast tomosynthesis. Medical Physics. 2014; 41:041913. [PubMed: 24694144]

Xu N, Yi S, Mendonca P, Tian T-p, Samala R, Chan H-P. False positive reduction of microcalcification cluster detection in digital breast tomosynthesis. Proc. SPIE. 2014; 9034:90342N.

Zhang Y, Chan H-P, Sahiner B, Wei J, Goodsitt MM, Hadjiiski LM, Ge J, Zhou C. A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis. Medical Physics. 2006; 33:3781–3795. [PubMed: 17089843]

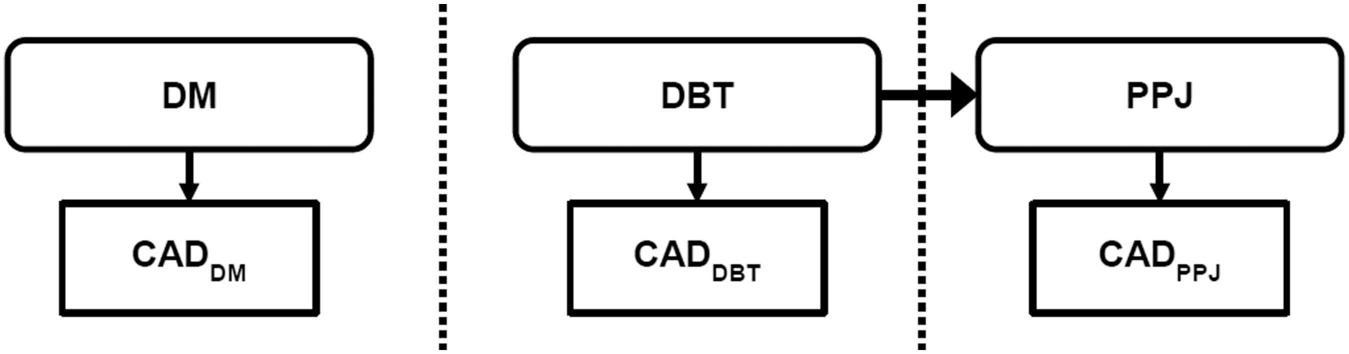


Fig. 1.
Overview of the CADe systems comparison. DM: digital mammography, DBT: digital breast tomosynthesis, PPJ: planar projection image. PPJ image is generated from DBT volume. All three CAD systems were tested on corresponding patient cases.

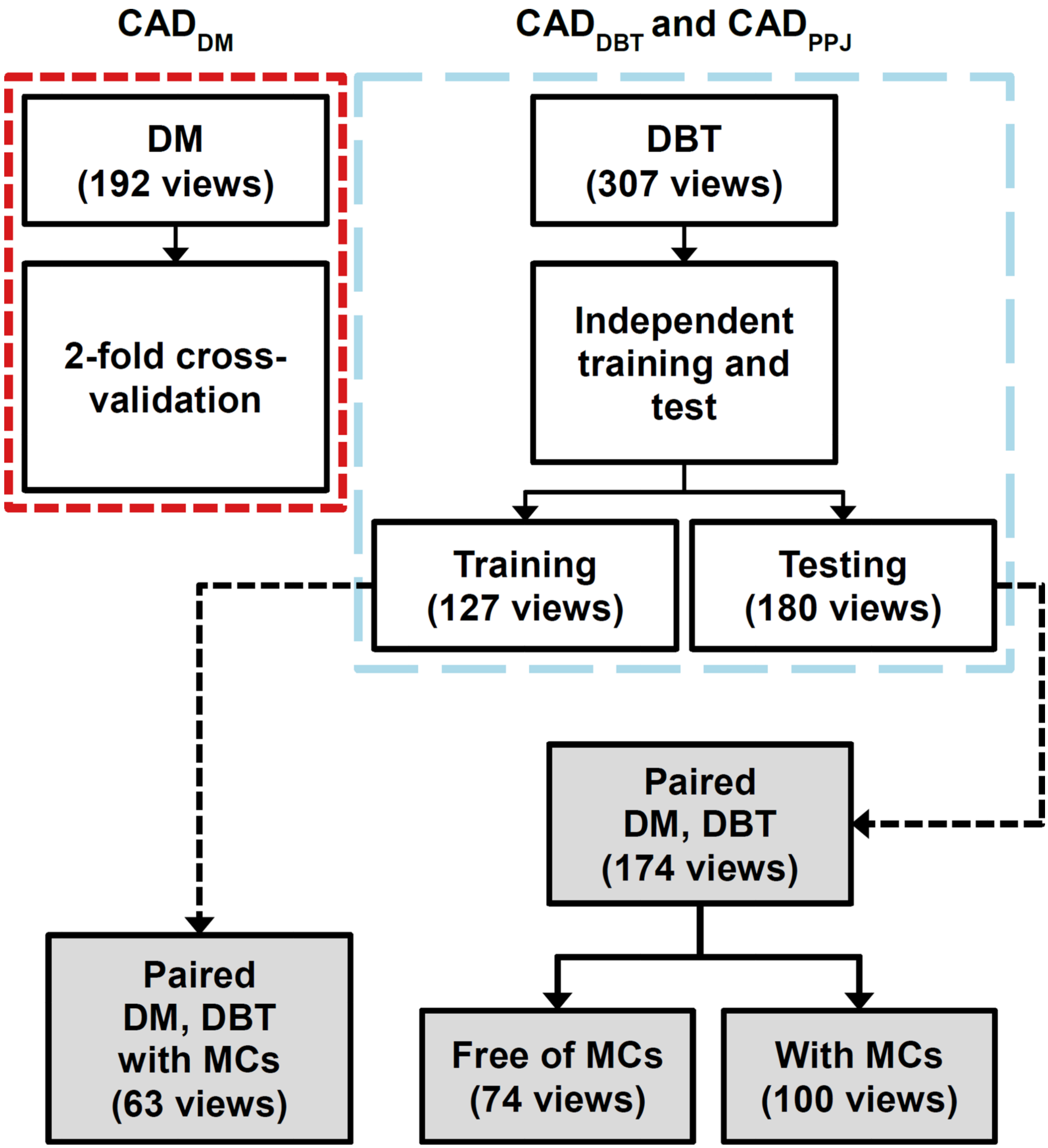


Fig. 2.

Data sets for DM, DBT and PPJ CADe systems. DBT and PPJ CADe systems use the same data set. Data sets marked in gray boxes were used in this study. CAD_{DM} was trained and validated on an independent data set of 192 views in previous studies.

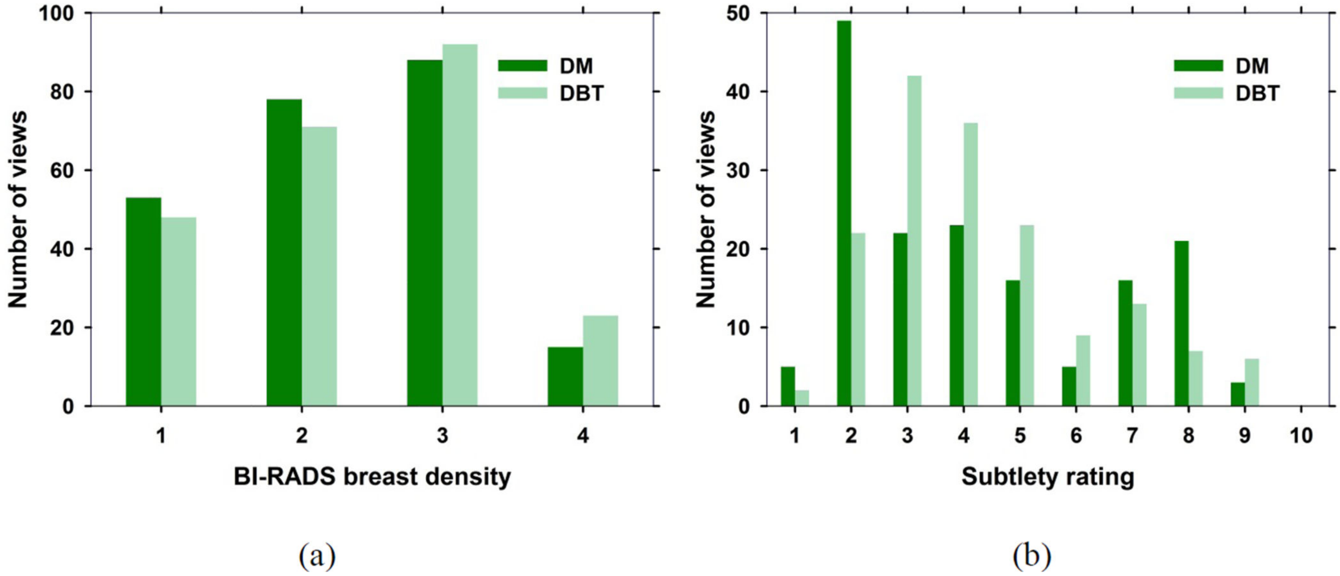


Fig. 3.

(a) BI-RADS breast density rating for the 237 views from the training and test set with and without MCs. (b) Subtlety rating for the MC clusters on 163 views from the training and test set with MCs. A subtlety rating of 1 refers to most visible cluster.

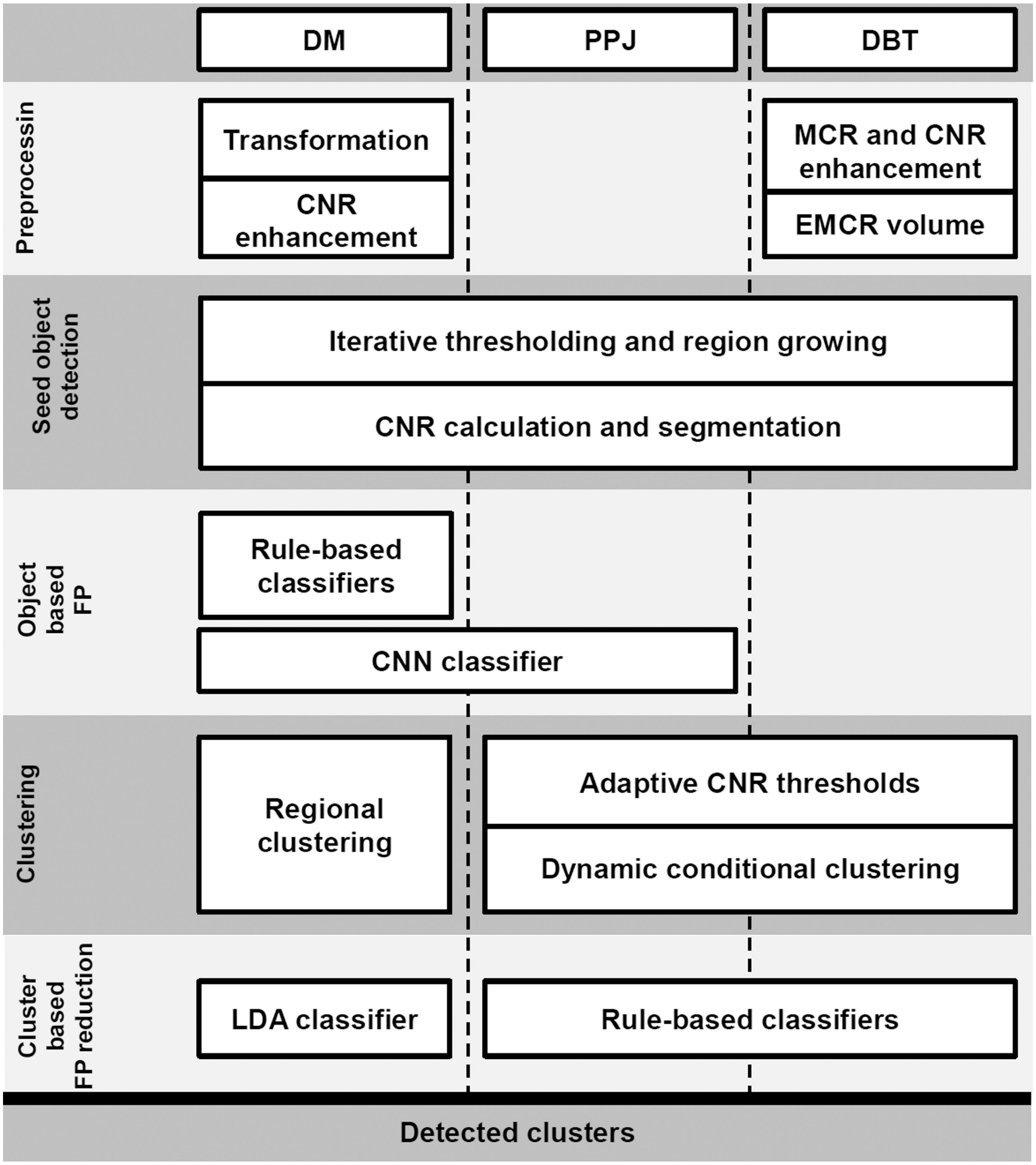


Fig. 4.

Flow diagram comparing different blocks of CADe systems between DM, DBT and PPJ. The dotted line separates the CADe systems and the blocks crossing the dotted line show methods common to the CADe systems in the columns. For example, the ‘CNN classifier’ block is used in both DM and PPJ CADe systems. Note that each block is specifically trained for each type of images although the general techniques are similar. (CNN: convolution neural network, LDA: linear discriminant analysis, MCR: multiscale calcification response, CNR: contrast-to-noise ratio, EMCR: enhanced modulated calcification response)

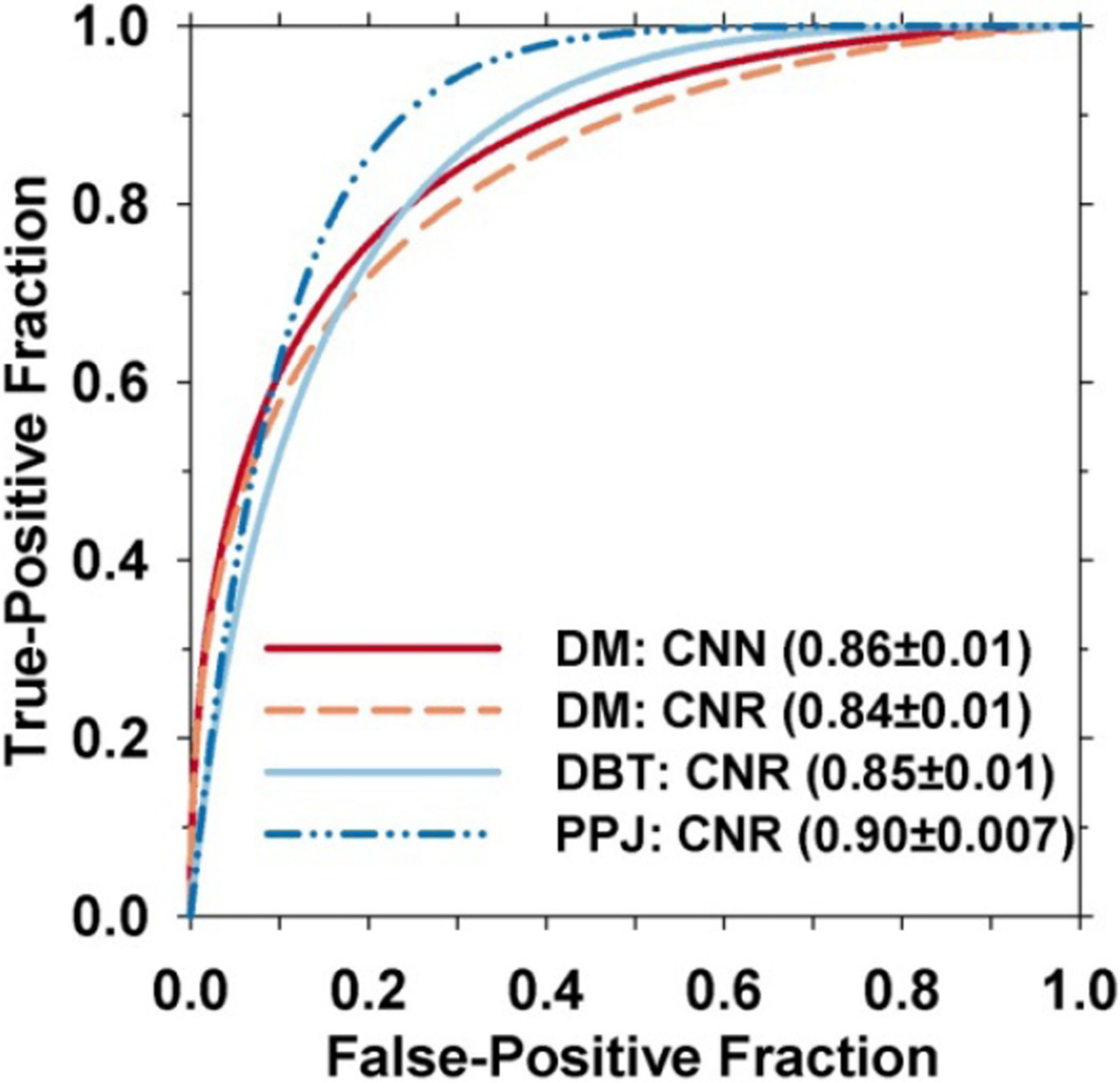


Fig. 5.

ROC curves for CNR_{obj} of corresponding objects between DBT and PPJ, and for CNN and CNR of objects for DM in the subset of 40 training views with manually marked microcalcifications. The CADe-detected candidates overlapped with manually marked objects are considered as TPs and the rest as FPs.

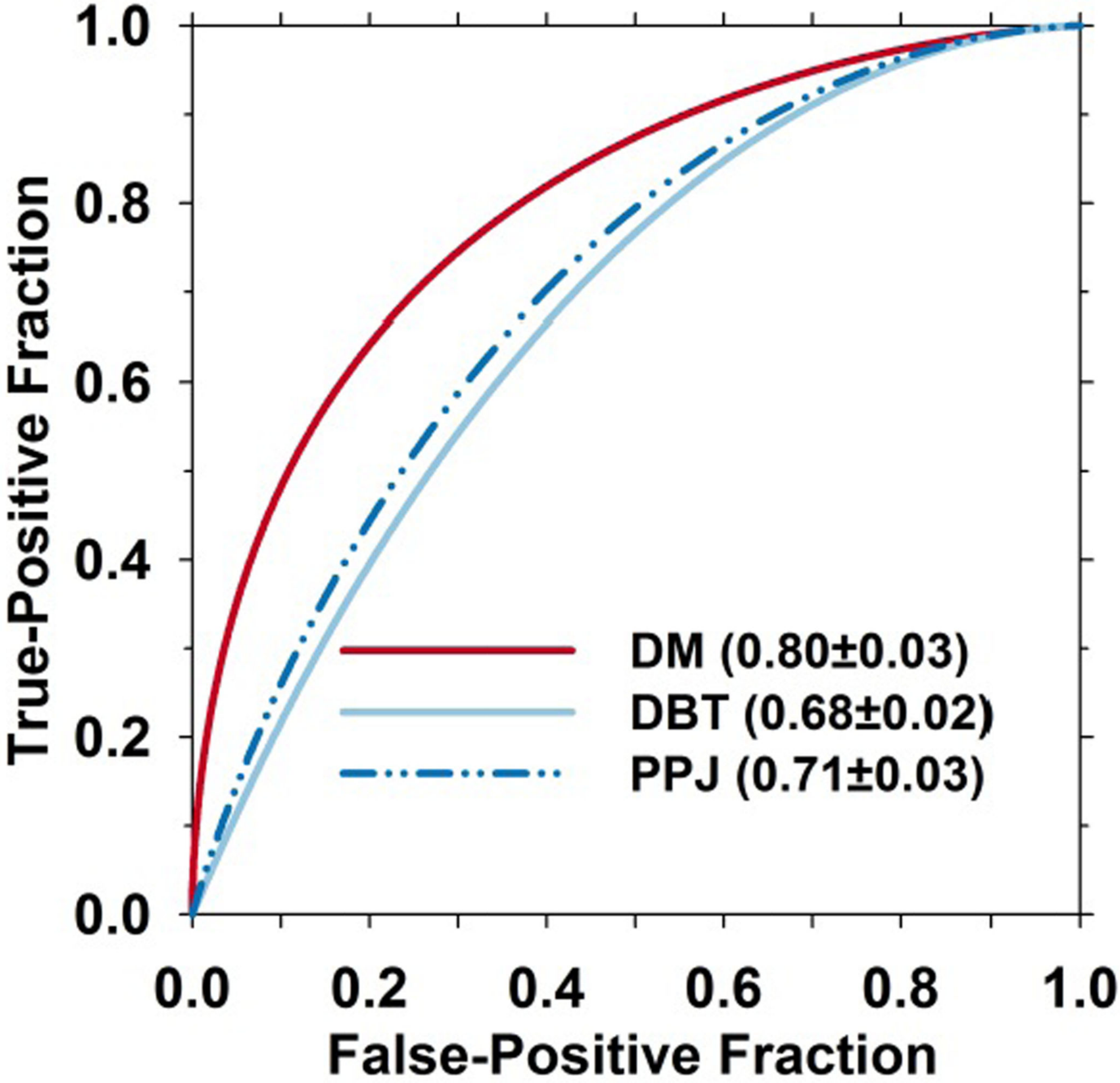


Fig. 6.

ROC curves of one of the features, number of candidates in a cluster, for the test set. The performances of the feature for the detected clusters in DM, DBT and PPJ are compared.

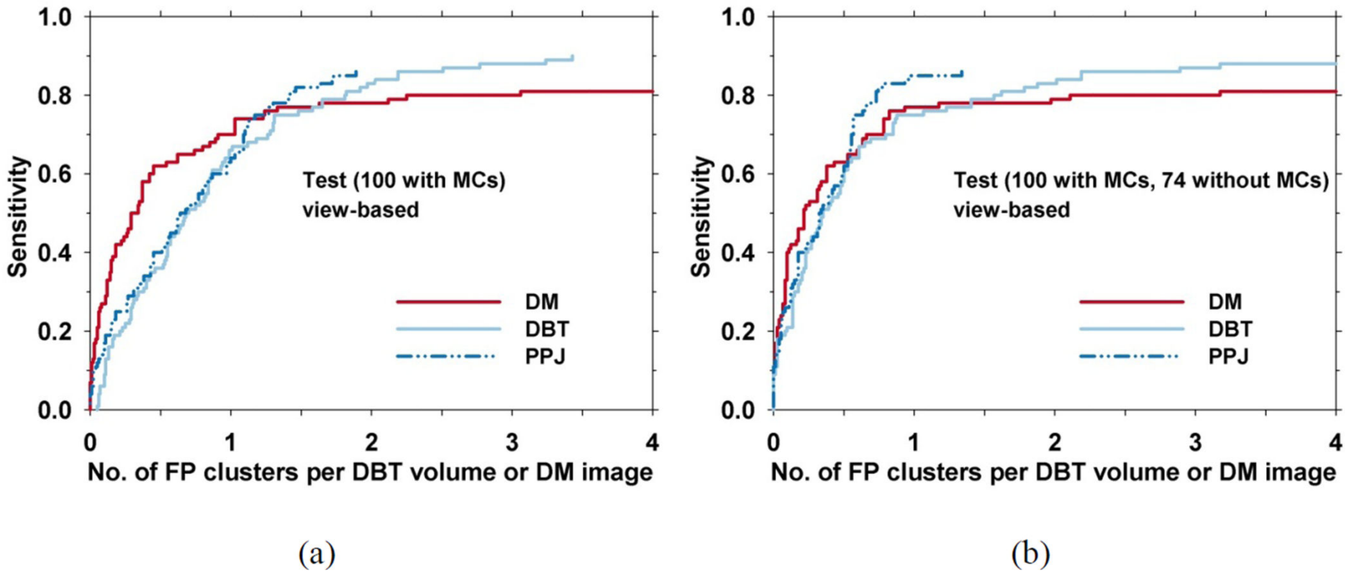
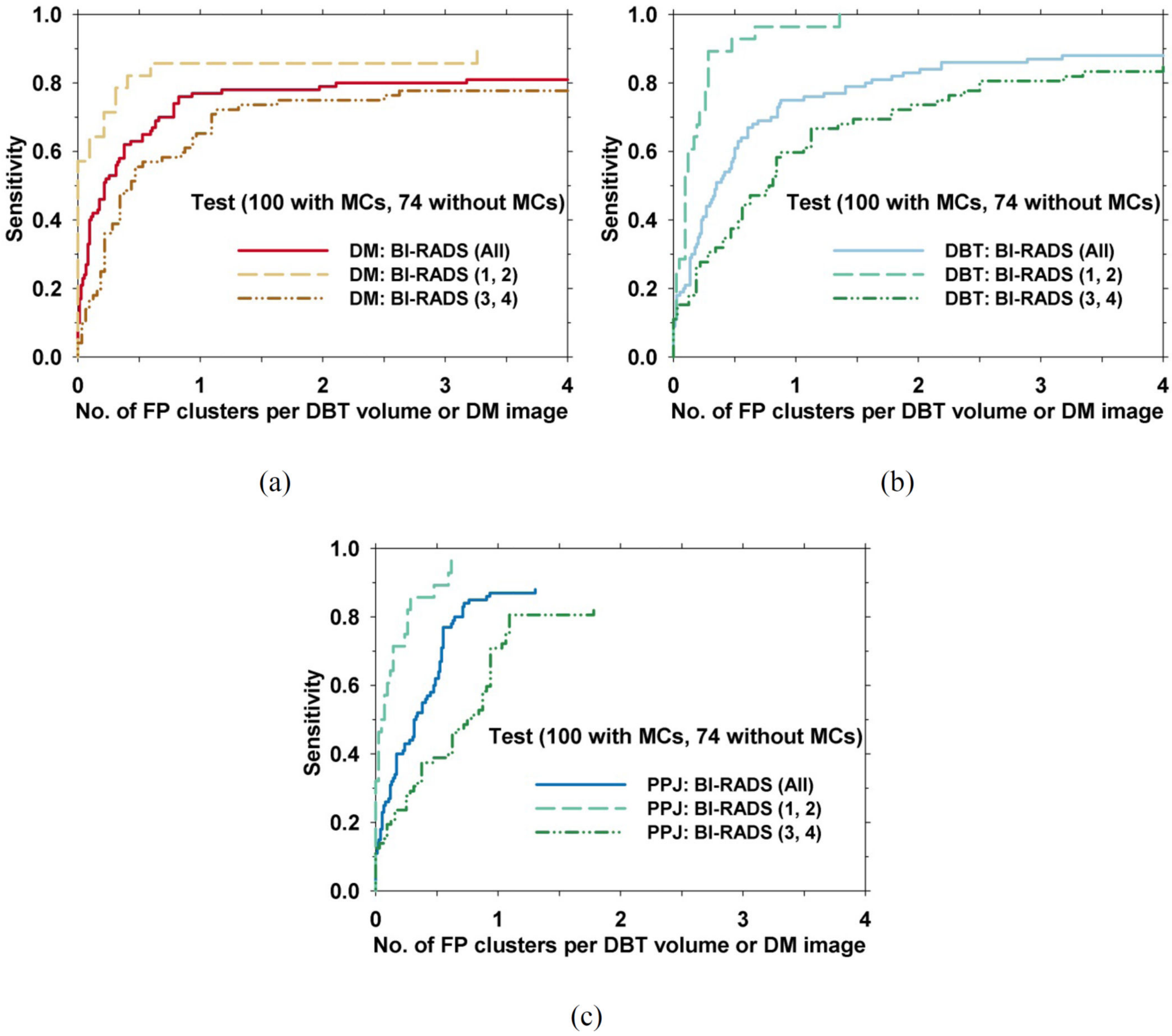


Fig. 7.

View-based FROC curves for the test set. (a) The average FP rate was estimated from views with MCs. (b) The average FP rate was estimated from views without MCs.



(a) (b)

(c)

Fig. 8.

Effect of BI-RADS density on the detection of MCs in the test set with the average FP rate estimated from views without MCs. FROC curves of the test set for the two density categories for (a) CAD_{DM}, (b) CAD_{DBT} and (c) CAD_{PPJ} systems.

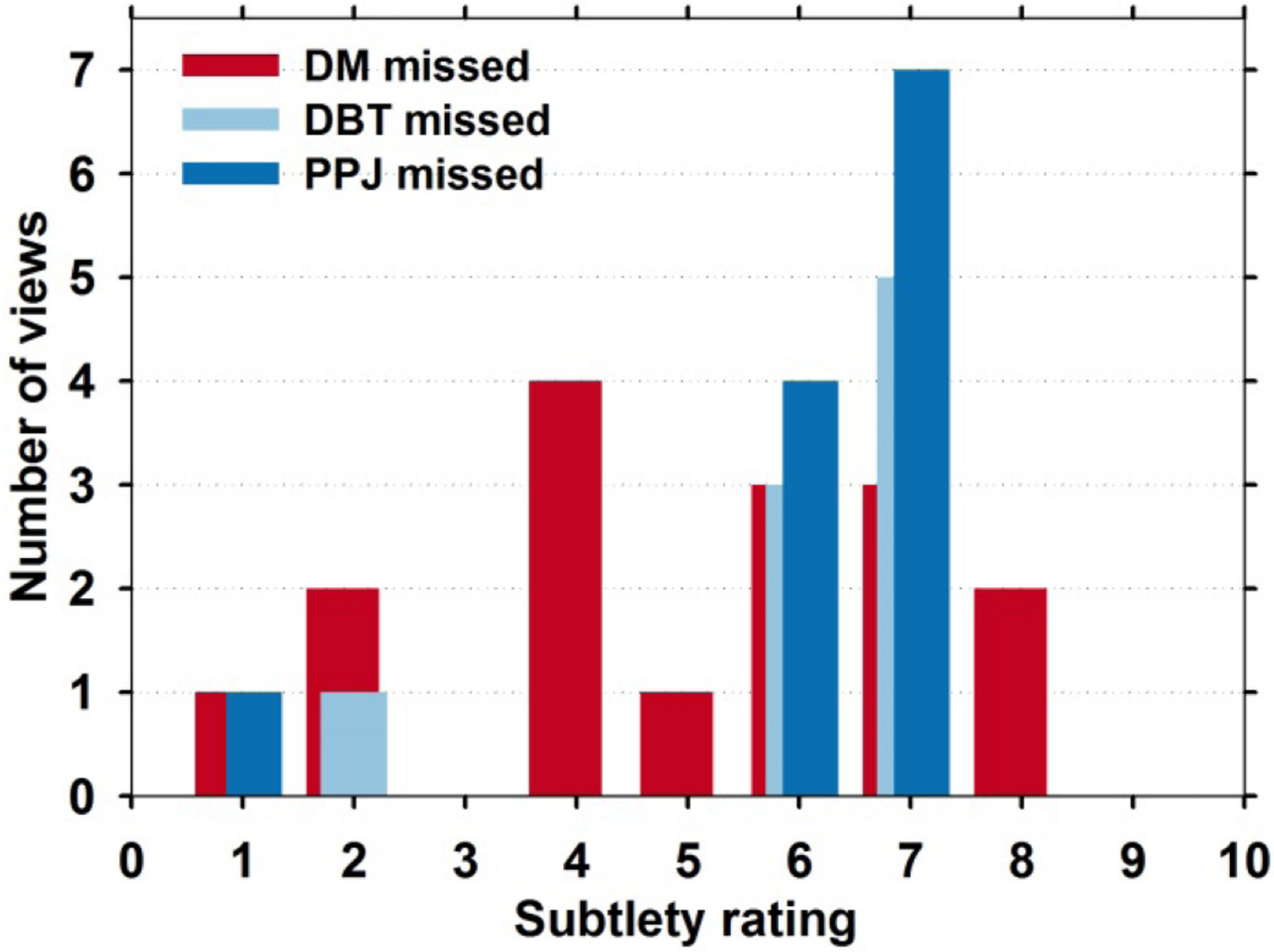


Fig. 9.

Histogram of subtlety ratings of all the false negative MCs by the CADe systems. CAD_{DM} alone lost 16 MCs, CAD_{DBT} alone lost 9 MCs and CAD_{PPJ} lost 12 MCs. Four of the lost MCs were missed by all systems.

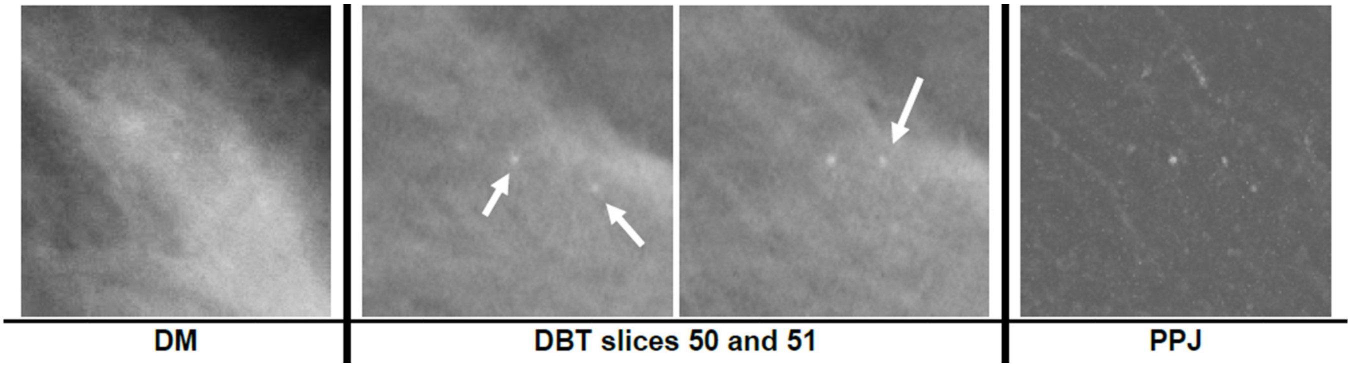


Fig. 10.

A region of interest of size 15 × 15 mm containing a cluster of ductal carcinoma in situ (DCIS) on DM and in DBT slices 50 to 51, which were at depths of about 50 mm to 51 mm from the compression paddle. The subtlety rating was 9 and 5 for DM and DBT, respectively (a rating of 1 indicates most visible). The three microcalcifications on the DBT slices are not clearly visible on the DM image. The object at the center of slices 50 and 51 is the same microcalcification.

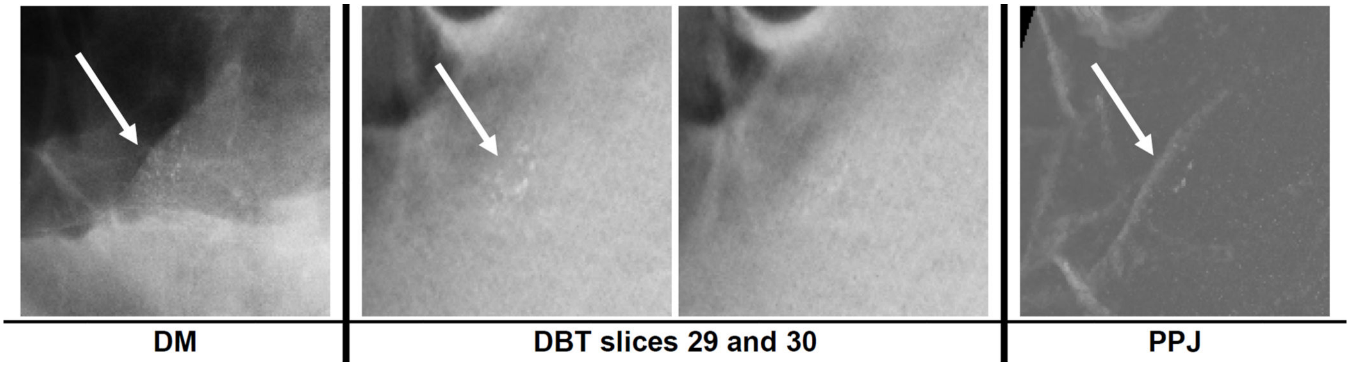


Fig. 11.

A region of interest of size 15 × 15 mm containing a benign microcalcification cluster on DM and in DBT slices 29 to 30, which were at depths of about 29 mm to 30 mm from the compression paddle. The subtlety rating was 4 and 7 for DM and DBT, respectively (a rating of 1 indicates most visible) with BI-RADS density rating of 3.

Table I

Literature review of CADe for MCs on DBT. The view-based sensitivities and FP rates are summarized.

Author, year	**Type**	**Properties**	**Data size and results**
(Peters *et al.*, 2006)	PV	For each PV image: multiscale wavelets for prescreening of candidates, calculation of membership values from fuzzy contours based on multilevel threshold, segmentation based on fuzzy particle map, aggregation of different objects and mapping to 3D locations.	Detection of microcalcifications only, not MCs. No quantitative results.
(Park *et al.*, 2008)	PV and RV	For each PV image or RV slice: threshold on the output of Difference-of-Gaussian (DoG) for prescreening of objects, multi-layer topographic region growing, labeling and clustering to generate potential MCs, artificial neural network for classification and generation of scores for cluster, grouping methods to match clusters or identify unique clusters.	V-96 *For PV:* 3.99 FPR at 70% S *For RV:* 15.85 FPR at 88% S
(Reiser *et al.*, 2008)	PV	For each PV image: morphologic denoising followed by DoG for detection of microcalcifications candidates, segmentation using local thresholding and backprojected to 3D feature volume to reduce FPs, reprojected back to PV images followed by linear discriminant analysis (LDA) and round robin method for MC detection.	V-60 (M-21, B-9, N-30) 1.3 FPR at 86% S
(Bernard *et al.*, 2008)	RV	Measure contrast of microcalcifications on each slice using Mexican hat wavelet responses followed by FP reduction methods based on difference in attenuation coefficients of microcalcifications and breast tissue and also based on background noise estimation, 3D connected component analysis for segmentation and 3D clustering for MC formation.	V-50 (N-37) 1.4 FPR at 85% S
(van Schie and Karssemeijer, 2009)	RV	Generate a noise model for iterative reconstruction algorithm with MC detection as an evaluation method. Uses a simple local contrast feature for segmentation and clustering using a 5 mm radius sphere.	V-41 (N-29) 1.5 FPR at 95% S
(Chan *et al.*, 2010)	RV	Two parallel processes involving 3D multiscale filtering and signal-to-noise ratio enhancement filter combined to generate enhancement-modulated calcification response volume, followed by detection of microcalcifications and 3D dynamic clustering results in MCs, MCs screened using cluster feature analysis to reduce FPs.	V-78 (M-33, B-41) 1.5 FPR at 82% S
(Sahiner *et al.*, 2012)	RV	DBT volume is enhanced based on contrast and structure, global thresholding and region growing followed by local thresholding results in potential microcalcification candidates, dynamic clustering to form MCs and rule-based classifiers for FP reduction.	V-220 (M-34, B-110, N-76) 3.4 FPR at 85% S
(Samala *et al.*, 2014b)	RV	Similar to Sahiner *et al* (Sahiner *et al.*, 2012), but DBT volume enhanced using multiscale bilateral filter regularization during reconstruction, weighted combination of contrast and structure enhanced volumes, thresholds for cluster formation are adaptively generated based on DBT volume, additional maximum-intensity-based FP reduction method.	V-384 (M-68, B-240, N-76) 2.16 FPR at 85% S
(Wei *et al.*, 2014)	PV	Each PV slice enhanced using CNR enhancement filter, iterative thresholding to detect initial MC candidates, decompose candidates using multichannel response analysis, the 2D response is backprojected into the 3D volume based on geometry of the DBT system, coincidence counting method for FP reduction, dynamic clustering results in MCs.	V-82 (M-32, B-10) 1.55 FPR at 90% S
(Samala *et al.*, 2014a)	PPJ	Introduction of PPJ image, a new representation of DBT volume generated from bilateral filtered high-frequency band extracted from regularized SART. High CNR objects screened through a convolutional neural network, clusters formed through dynamic conditional clustering followed by	V-307 (M-67, B-157, N-76) 0.71 FPR at 85% S

Author, year	**Type**	**Properties**	**Data size and results**
		feature-based FP reduction.	
(Xu *et al.*, 2014)	RV	Uses microcalcification candidates from Sahiner *et al* (Sahiner *et al.*, 2012), clusters formed by iterative clustering are classified using random-forest and rule-based classifier, FP reduction achieved through combination of the two classifier outputs.	V-80 0.5 FPR at 92% S
(Samala *et al.*, 2015)	RV and PPJ	Combines the detection in RV (Samala *et al.*, 2014b) and PPJ (Samala *et al.*, 2014a) for a narrow tomographic angle DBT system. Combination of RV- and PPJ-specific detection strategies to improve the detection sensitivity of the joint-CAD system.	V-307 (M-67, B-157, N-76) 1.72 FPR at 85% S

PV – projection view, RV – reconstructed volume, PPJ – planar projection image, V – number of views, M – number of malignant MCs, B – number of benign MCs, N – number of views without MCs, FPR – number of false positives/DBT view, S – sensitivity.

Table II

Non-parametric analysis of the significance of the differences in MC detection performance of the three CADe systems, $\mathrm{CAD}_{\mathrm{DBT}}$, $\mathrm{CAD}_{\mathrm{PPJ}}$ and $\mathrm{CAD}_{\mathrm{DM}}$, for the test set where the FP rates were estimated from views with MCs. The operating threshold was set at 1 FP/view for the analysis. The figure-of-merit (FOM) is the difference in the areas under the two FROC curves being compared, CI: 95% confidence interval.

CADe	**FOM**	**CI**	***p* value**
DM DBT	0.169 (DM > DBT)	(0.05, 0.29)	0.013*
DBT PPJ	0.010 (PPJ > DBT)	(−0.08, 0.1)	0.847
DM PPJ	0.159 (DM > PPJ)	(0.04, 0.27)	0.006*

Table III

JAFROC analysis of the significance of the differences in MC detection performance for three CADe systems, $\mathrm{CAD}_{\mathrm{DBT}}$, $\mathrm{CAD}_{\mathrm{PPJ}}$ and $\mathrm{CAD}_{\mathrm{DM}}$, on the test set, where the FP rates were estimated from views without MCs. FOM: figure-of-merit, CI: 95% confidence interval.

CADe	FOM	CI	*p* value
DM	0.67	(0.58, 0.75)	0.924
DBT	0.67	(0.59, 0.75)	
DBT	0.67	(0.59, 0.75)	0.034*
PPJ	0.74	(0.67, 0.82)	
DM	0.67	(0.58, 0.75)	0.059
PPJ	0.74	(0.67, 0.82)	