

A Comparative Analysis of 2D and 3D CAD for Calcifications in Digital Breast Tomosynthesis

Raymond J. Acciavatti, Shonket Ray, Brad M. Keller,
Andrew D. A. Maidment, and Emily F. Conant

University of Pennsylvania, Department of Radiology, 3400 Spruce St., Philadelphia PA 19104

E-mail: {Raymond.Acciavatti | Shonket.Ray | Brad.Keller |
Andrew.Maidment | Emily.Conant}@uphs.upenn.edu

ABSTRACT

Many medical centers offer digital breast tomosynthesis (DBT) and 2D digital mammography acquired under the same compression (*i.e.*, “Combo” examination) for screening. This paper compares a conventional 2D CAD algorithm (Hologic® ImageChecker® CAD v9.4) for calcification detection against a prototype 3D algorithm (Hologic® ImageChecker® 3D Calc CAD v1.0). Due to the newness of DBT, the development of this 3D CAD algorithm is ongoing, and it is currently not FDA-approved in the United States. For this study, DBT screening cases with suspicious calcifications were identified retrospectively at the University of Pennsylvania. An expert radiologist (E.F.C.) reviewed images with both 2D and DBT CAD marks, and compared the marks to biopsy results. Control cases with one-year negative follow-up were also studied; these cases either possess clearly benign calcifications or lacked calcifications. To allow the user to alter the sensitivity for cancer detection, an operating point is assigned to each CAD mark. As expected from conventional 2D CAD, increasing the operating point in 3D CAD increases sensitivity and reduces specificity. Additionally, we showed that some cancers are occult to 2D CAD at all operating points. By contrast, 3D CAD allows for detection of some cancers that are missed on 2D CAD. We also demonstrated that some non-cancerous CAD marks in 3D are not present at analogous locations in the 2D image. Hence, there are additional marks when using both 2D and 3D CAD in combination, leading to lower specificity than with conventional 2D CAD alone.

Keywords: Computer-aided detection (CAD), digital breast tomosynthesis (DBT), digital mammography (DM), calcifications, breast cancer, technology evaluation, sensitivity, specificity.

1. INTRODUCTION

Although 2D digital mammography (DM) has helped to promote early detection of breast cancer, one drawback of this screening technique is tissue superposition. A cancerous mass can be hidden by normal structures that project onto a similar position in the image. Many medical centers now offer digital breast tomosynthesis (DBT) as an adjunct or alternative to DM for breast cancer screening exams. In DBT, a 3D image is created from multiple 2D x-ray projection images acquired over a small range of angles; for example, 15° for the Selenia Dimensions system (Hologic, Inc., Bedford, MA, USA).¹

The University of Pennsylvania (Penn) was one of the first large medical centers to offer DBT and DM acquired under the same compression (*i.e.*, “Combo Mode”) for all screening exams (since September of 2011). Based on Penn’s experience with acquiring screening images in Combo Mode, it was determined that DBT reduces the recall rate and increases the detection rate of cancers compared to screening with DM alone.² To minimize the total radiation dose, Penn recently implemented C-View™ as an alternate method for creating the 2D image. The C-View™ image is created directly from the 3D data set.³

Despite these benefits, one potential hurdle for DBT is calcification imaging. Multiple studies have suggested that calcification clusters are not as clearly visible in DBT.⁴⁻⁸ To determine whether the benefits of DBT are limited to noncalcification findings, Rafferty *et al.* performed reader studies using receiver operating characteristic (ROC) methods.⁹ The diagnostic accuracy of Combo Mode was found to be superior to DM alone; however, the gain in reader performance was attributed entirely to the noncalcification findings. There was no statistically significant difference between Combo Mode and DM alone in the ROC curves for calcifications findings.

Recently, a prototype computer-aided detection (CAD) algorithm for calcification detection in DBT has been introduced by Hologic, the company that manufactures the imaging systems used for screening exams at Penn. This paper evaluates the potential merits of this prototype CAD algorithm (Hologic® ImageChecker® 3D Calc CAD v1.0).³ Unlike the conventional 2D CAD algorithm for DM (Hologic® ImageChecker® CAD v9.4), the 3D CAD algorithm is not yet approved by the United States Food and Drug Administration (FDA), and is intended for investigational use in the United States. Drawing from Penn’s database of screening exams, the cases considered in this project span a broad range of features, including:

- i. Screening cases with suspicious calcifications – proven by biopsy to be benign, high risk, or malignant;
- ii. Screening control cases with an overall assessment of 1 (negative) or 2 (benign) using the Breast Imaging-Reporting and Data System (BI-RADS®) – validated with follow-up imaging studies.

The purpose of this study is to compare the accuracy of 2D and 3D CAD in terms of the detection of clinically significant calcifications (biopsy-proven malignancy), since “truth” is known for each case either by biopsy results or by follow-up imaging studies.

2. METHODS

2.1 Description of the Data Set

In this study, 51 DBT screening cases for which stereotactic breast biopsy was recommended for suspicious calcifications were analyzed. Cases were identified retrospectively at the Hospital of the University of Pennsylvania (HUP) between May of 2012 and August of 2014. The data collection was approved by the Institutional Review Board (IRB) at Penn and was compliant with the Health Insurance Portability and Accountability Act (HIPAA). Stereotactic biopsies that were performed for reasons other than suspicious calcifications were excluded from the data set; for example, focal asymmetries.

As shown in Figure 1, one patient was recommended for a bilateral biopsy, and as a result, there were biopsy results for 52 individual breasts. Table 1 provides a summary of the histological findings in the biopsied breasts. Out of 25 breasts that were found to be “positive” for cancer, there were 19 breasts with ductal carcinoma *in situ* (DCIS). Additionally, there were six breasts with invasive cancer. One would expect fewer invasive cancers than *in situ* cancers, as was observed in this study, since biopsies were performed on the basis of suspicious calcifications alone. Most invasive cancers are associated with a mass, and hence are biopsied with ultrasound guidance needed; stereotactic biopsy is usually reserved for calcium-only lesions.

Breasts with a non-cancerous biopsy were considered “negatives”. Out of 27 breasts with a negative biopsy, 25 breasts had benign calcifications and two breasts had high-risk calcifications (Table 1). The histological finding for both high-risk cases was atypical ductal hyperplasia.

In order to enhance the number of “negatives” that are considered in the data set, one can also analyze the breast that is contralateral to the biopsy site. Follow-up studies were used to determine whether this breast could truly be considered “negative”. All follow-up imaging studies at HUP were considered for this purpose, including x-ray exams, ultrasound, and magnetic resonance imaging (MRI) obtained approximately one year after the study case. Variations in the follow-up time for each patient are quantified in Section 3.2. Due to lack of follow-up, it was necessary to exclude 12 breasts that were contralateral to the biopsy site (Figure 1). In addition, it was necessary to exclude two breasts possessing cancer that was predominately a mass. An analysis of mass detection is beyond the scope of this study, since Hologic® ImageChecker® 3D Calc CAD v1.0 does not include mass marks. After these exclusions were performed, it was found that 36 of the breasts that were contralateral to the biopsy site could be considered “negatives” for the purpose of this study.

Table 1. The biopsy (“Bx”) findings for calcifications (“calcs”) are summarized below.

	Bx Results Studied	Benign Calcs on Bx	High-Risk Calcs on Bx	<i>In Situ</i> Ductal Cancer on Bx	Invasive Cancer on Bx
Non-Cancer	27 (51.9%)	25 (48.1%)	2 (3.8%)	---	---
Cancer	25 (48.1%)	---	---	19 (36.5%)	6 (11.5%)

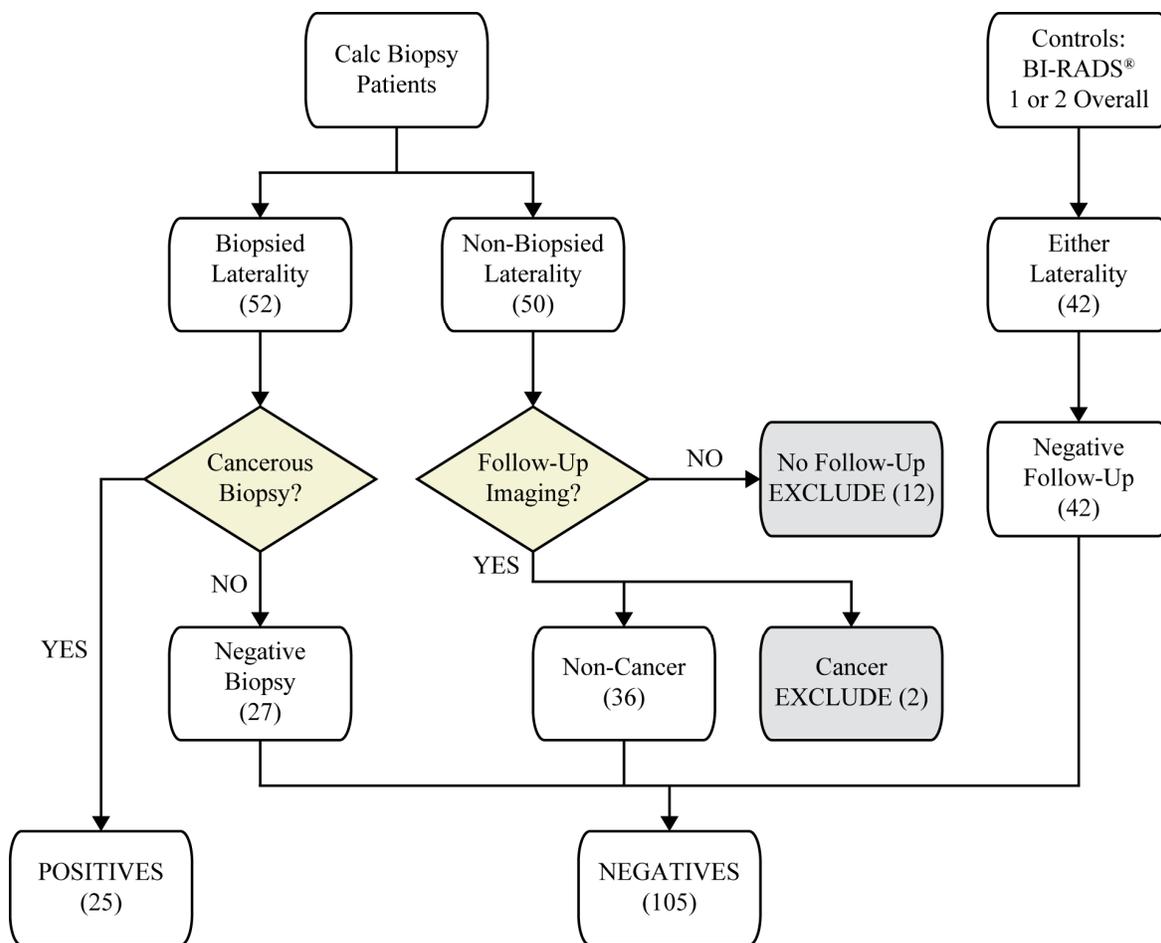


Figure 1. As shown by the flow diagram, breasts are considered “positive” based on the presence of a biopsy-proven malignancy. Three distinct strata are identified for “negative” breasts; in particular, *i.* breasts possessing a negative biopsy, *ii.* breasts that are contralateral to the biopsy site, and *iii.* control breasts.

To enhance the number of “negatives” further in this data set, one can consider an entirely different group of patients that were not recommended for stereotactic biopsy. Screening exams from an additional 21 control patients were identified retrospectively between May and October of 2012 for this purpose. Their overall BI-RADS[®] assessment was either category 1 (negative) or category 2 (benign). To verify that cancer was not detected over time in these patients, all follow-up imaging studies at HUP were analyzed. All 21 control patients were found to have an overall BI-RADS[®] assessment of 1 or 2 in follow-up studies of at least one-year minimum. Since two lateralities (left and right) were imaged per case, 42 individual breasts were analyzed with these characteristics (Figure 1).

2.2 CAD Scoring Rules

In the 2D DM image, three types of CAD marks are possible: *i.* triangle (calcification), *ii.* star (mass), and *iii.* cross (calcification + mass). By contrast, only one type of CAD mark can be observed in the 3D DBT image; namely, the triangle (calcification). For the purpose of this study, only triangle and cross marks on 2D DM images and only triangle marks on 3D DBT images were analyzed. This convention was chosen since this study is designed to evaluate CAD in the task of calcification detection. An analysis of mass detection is beyond the scope of this work.

An expert radiologist (E.F.C.) reviewed the individual images and pathology reports to determine the location of the biopsy site. The pathology reports indicated whether the biopsy site was benign (normal), high-risk (atypia), or malignant (cancerous). A CAD mark including any position on a biopsy-proven malignancy was considered a true positive (TP). All other CAD marks in the image were considered false positive (FP). In particular, a CAD mark on a lesion that was proven by biopsy to be benign or high-risk was considered a FP. Additionally, a CAD mark that was positioned on a non-biopsied location in the image was deemed a FP.

In some patients, it is not possible to image the entire breast in a single view due to the breast size. As a result, additional views are acquired after re-positioning the breast. To minimize the radiation dose in a screening exam, the additional views are typically acquired with 2D DM imaging alone. The view in Combo Mode (3D DBT + 2D DM) is reserved for imaging the breast area with the maximum glandularity. For the purpose of comparing 2D and 3D CAD, all views that were not acquired in Combo Mode were not scored. Consequently, each 2D image in this study is associated with a 3D image having the same positioning and vice versa.

In the Extensible Markup Language (XML) report, each CAD mark is assigned a certainty ranging between zero and unity. Table 2 shows that CAD marks are in turn assigned an operating point based on the certainty. In particular, CAD marks with very low certainty are assigned an operating point of “-1”. The certainty range for operating point “-1” is between 0 and 0.41 for 2D CAD and is between 0 and 0.27 for 3D CAD. Although CAD marks at operating point “-1” were calculated in the XML report, these CAD marks were not displayed on the SecurView® 8.1 Workstation that was used to view the images. For this reason, it was not possible for the radiologist to score these CAD marks as either TP or FP. CAD marks at all other operating points (0, 1, and 2) were displayed on the SecurView® 8.1 Workstation, and therefore, were visible to the radiologist for scoring.

Unlike 3D CAD which includes only triangle marks, 2D CAD includes an additional mark denoting the presence of both a calcification cluster and a mass (*i.e.*, a cross mark). It was found that the certainty assigned to the calcification component of a cross mark is not necessarily equivalent to the certainty assigned to the mass component of the same mark. For the purpose of analyzing calcification detection, only the certainty corresponding to the calcification component of a cross mark was considered in this study.

Table 2. An operating point is assigned to each CAD mark based on the certainty.

2D CAD	Operating Point	Certainty Threshold	3D CAD	Operating Point	Certainty Threshold
	0	0.72		0	0.81
	1	0.50		1	0.55
	2	0.41		2	0.27
	-1 (Not Scored)	<0.41		-1 (Not Scored)	<0.27

2.3 Statistical Calculations

To calculate the sensitivity of CAD, one must determine whether the cancer was detected in breasts possessing a biopsy-proven malignancy. For the purpose of this work, the existence of at least one TP in an image was considered a lesion localization (LL). An image with at least one TP was assigned a LL value of unity, irrespective of the total number of TPs and FPs in the image. Images lacking a TP were assigned a LL value of zero. Under this convention, the Lesion Localization Fraction (LLF) was calculated as the ratio of LLs to the number of biopsy-proven malignancies (25).

In addition, one can calculate the total number of FPs in all images, and then divide the result by the total number of images considered. This calculation yields the mean number of FPs per image, or the Non-Lesion Localization Fraction (NLF). Whereas the LLF is bounded between zero and unity, the NLF can be arbitrarily large depending on the number of FPs in the images.¹⁰

3. RESULTS

3.1 Quantifying the Sensitivity of CAD

To analyze the sensitivity of CAD, the LLF is plotted as a function of certainty in Figure 2(a). As one moves to the right in this plot, the certainty threshold is increased, and hence TPs with lower certainty than the threshold are not counted. The LLF is a decreasing function of certainty for this reason. The plot demonstrates that 2D CAD has a modest improvement in sensitivity over 3D CAD in the CC view. At a certainty of 0.41, the LLF is 1.00 for 2D CAD and is 0.96 for 3D CAD. The modest sensitivity advantage of 2D CAD is preserved at additional certainties. Although marks

with certainty less than 0.41 were not visible on the SecurView® 8.1 Workstation in the 2D images, marks with lower certainty were indeed visible in the 3D images. Figure 2(a) demonstrates that there was no improvement in the sensitivity of 3D CAD as the certainty threshold was reduced from 0.41 to 0.27.

Although all cancers in the CC view were detected by 2D CAD, there was one cancer that was occult to 3D CAD in the CC view. This case is illustrated in Figure 3. The slice in the 3D image for which the cancer is in focus is shown in Figure 3(a). The cancer, which is less than 1.0 cm in size, is denoted by the arrow. According to the pathology report, the histology of this cancer was high-grade DCIS with invasion. The cancer was not marked by 3D CAD, even though it is clearly visible. It should be emphasized that there was no mark in the XML report with an operating point of “-1”. If such a mark were present in the XML report, it would be possible that the cancer was detected by 3D CAD but simply not displayed on the SecurView® 8.1 Workstation that was used to visualize the CAD marks. An analysis of why this cancer was not detected by 3D CAD is beyond the scope of this work.

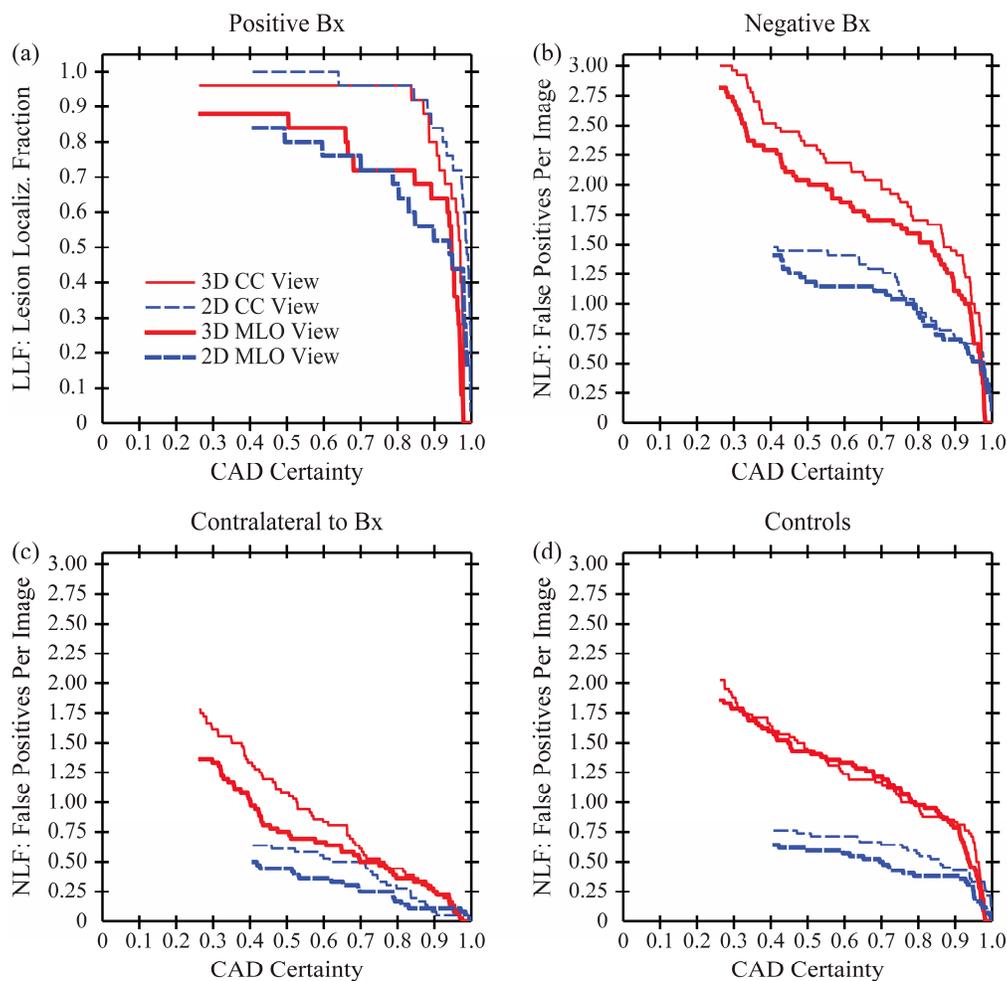


Figure 2. (a) Although 2D CAD exhibits a modest sensitivity advantage over 3D CAD in the CC view, the opposite result holds in the MLO view. (b) In breasts possessing a negative biopsy, the NLF is calculated. As shown, there are more FPs in the 3D image than in the 2D image. (c) Similarly, in breasts that are contralateral to the biopsy site, there are more FPs marked by 3D CAD than by 2D CAD. Since these breasts were not recommended for biopsy, one would expect there to be less FPs in comparison to breasts possessing a negative biopsy. (d) The NLF plot for control breasts provides additional evidence that 3D CAD is characterized by a larger number of FPs than 2D CAD. All subplots implicitly share a common legend.

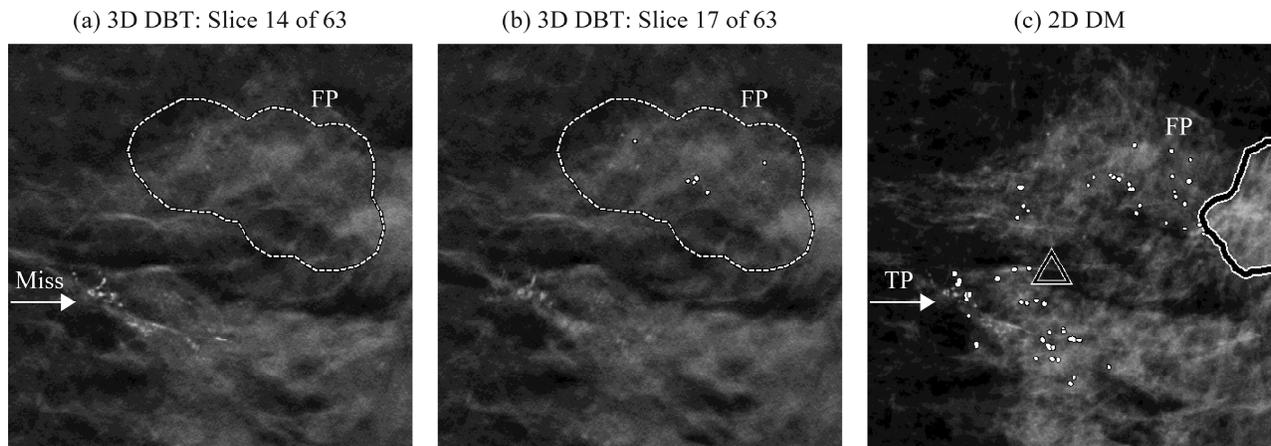


Figure 3. (a) In this CC view, a biopsy-proven malignancy (see arrow) is not marked by 3D CAD, illustrating a potential drawback of the technology. (b) Although a calcification cluster is marked by 3D CAD, none of the marked area includes the cancer. (c) Unlike the 3D CAD mark, the 2D CAD mark includes the cancer. It is noteworthy that the marked area overestimates the extent of the cancer, since non-cancerous calcifications are included. The mass mark in the right-hand side of the image was not analyzed for the purpose of this work.

Within a few centimeters of the cancer, a FP with a certainty of 0.913 was marked by 3D CAD [Figure 3(b)]. Although the FP spans slices 15 to 24, only slice 17 is shown for illustration. The cancer itself is not in focus in slice 17, and therefore appears blurry in comparison to slice 14 [Figure 3(a)]. Also, the dashed white outline of the FP is shown in Figure 3(a), even though there are no individual calcifications being marked in slice 14. For this reason, it was determined that the dashed white outline of the calcification cluster spans a broader thickness range than the individual calcifications in the cluster.

Although the cancer is missed by 3D CAD, the cancer is indeed evident using 2D CAD. As shown in Figure 3(c), a broader area is marked in the 2D image than in the 3D image. The 2D mark includes both cancerous and non-cancerous elements, and has a higher certainty (0.998) than the 3D CAD mark. A small portion of a mass mark is shown in the right-hand side of Figure 3(c), but this mark was not analyzed based on reasons described in Section 2.

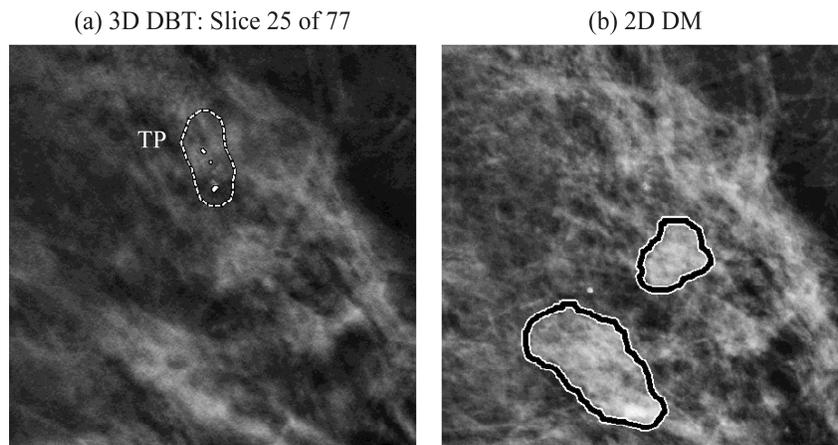


Figure 4. (a) In this MLO view, the cancer is detected perfectly by 3D CAD. (b) In this 2D DM image, dense tissue is superimposed over the calcification cluster, potentially obscuring its visibility. The cancer is not detected by 2D CAD. It should be emphasized that the two mass marks in the 2D DM image were not scored, since an analysis of mass detection is beyond the scope of this work.

While 2D CAD has a modest sensitivity advantage over 3D CAD in the CC view, the opposite result holds in the MLO view [Figure 2(a)]. At a certainty of 0.41, the LLF is 0.84 for 2D CAD and is 0.88 for 3D CAD. The modest sensitivity advantage of 3D CAD is preserved over a broad range of certainties. Although 3D marks with certainty less than 0.41 were visible on the SecurView[®] 8.1 Workstation, there was no change in sensitivity as the certainty threshold was reduced from 0.41 to 0.27.

To illustrate the sensitivity advantage of 3D CAD, Figure 4 shows a MLO view in which the cancer is detected by 3D CAD but not by 2D CAD. In this case, dense tissue is superimposed over the calcification cluster in the 2D image, potentially obscuring its visibility [Figure 4(b)]. Although marks with an operating point of “-1” are not displayed on the SecurView[®] 8.1 Workstation that was used to score the images, there was no such mark in the XML report. Consequently, it is not possible for the cancer to have been detected by 2D CAD at very low certainty (less than 0.41). Also, it should be emphasized that the two mass marks shown in the 2D image are not being analyzed for the purpose of this work.

Figure 4(a) demonstrates that the calcification cluster was indeed detected by 3D CAD (certainty 0.945). Unlike a single 2D image, a 3D image can filter out overlapping structures that may hide a calcification cluster. According to the pathology report, the histology of this cancer was DCIS. Although only one slice in the calcification cluster is shown (slice 25), additional calcifications were marked between slices 22 and 27.

3.2 Quantifying the False Positives of CAD

To quantify the mean number of FPs per image, the NLF was calculated [Figure 2(b)-(d)]. Similar to the LLF, the NLF is plotted as a function of certainty. As one moves to the right in this plot, the certainty threshold is increased, and hence fewer FPs are counted. The NLF was first determined for breasts possessing a negative biopsy. Figure 2(b) shows that at all certainties, the NLF is higher in the 3D image than in the 2D image, and thus the CAD performance is poorer in the 3D image. For example, at the lowest certainty that was scored in 2D CAD (0.41), the NLF is 1.48 for the 2D CC view and is 2.52 for the 3D CC view, giving a 70.3% increase in the mean number of FPs per image. The analogous values in the MLO view are 1.41 and 2.30, giving a 63.1% increase in the mean number of FPs per image. As shown, additional CAD marks with certainty below 0.41 were scored in the 3D image but not in the 2D image. Figure 2(b) demonstrates that the NLF of 3D CAD increases going from a certainty of 0.41 to 0.27.

The NLF was also quantified in breasts that were contralateral to the biopsy site [Figure 2(c)]. Although there were 50 individual breasts that were contralateral to the biopsy site, only 36 breasts were considered for this calculation based on the exclusion criteria described in Figure 1. One such criterion was the availability of follow-up studies. To quantify the time between the initial screening exam and the follow-up studies, a boxplot is shown in Figure 5(a). In patients with multiple follow-up studies, the longest time interval between the initial screening exam and follow-up was considered. Figure 5(a) shows that most patients had at least one-year follow-up (median 1.33 years). The 25th and 75th percentiles ranged between 1.03 and 2.00 years, giving an interquartile range (IQR) of 0.97 year. There were four patients out of 36 (11.1%) with slightly less than one-year follow-up, with the minimum follow-up time being 0.76 year. The remaining 32 patients had at least one-year follow-up. The maximum follow-up time was 2.23 years.

Figure 2(c) indicates that, in breasts contralateral to the biopsy site, the NLF was higher in the 3D image than in the 2D image. This result is similar to Figure 2(b) showing the NLF in breasts possessing a negative biopsy. At a certainty of 0.41, the NLF is 0.64 for the 2D CC view and is 1.31 for the 3D CC view, giving a 104.7% increase in the mean number of FPs per image. At the same certainty, the NLF is 0.50 for the 2D MLO view and is 0.97 for the 3D MLO view, giving a 94.0% increase in the mean number of FPs per image.

There is a strong difference in the NLF comparing breasts with a negative biopsy against breasts that are contralateral to the biopsy site. This difference can be quantified by subtracting the NLF in Figure 2(c) from the corresponding value in Figure 2(b). For example, at a certainty of 0.41, the change in NLF going from Figure 2(b) to Figure 2(c) is 0.84 (56.9% decrease), 0.91 (64.5% decrease), 1.21, (48.2% decrease), and 1.32 (57.7% decrease) for the 2D CC, 2D MLO, 3D CC, and 3D MLO views, respectively. One would expect the NLF to be larger in Figure 2(b) than in Figure 2(c), as was observed in this study. Since a breast that is recommended for biopsy is considered to be suspicious of cancer by a radiologist, breasts possessing a negative biopsy are inherently more likely to have FPs than breasts that are not recommended for biopsy.

In order to enrich the data set with an additional group of “negative” images, 21 control patients were also considered. A boxplot is shown in Figure 5(a) to quantify the time between the initial screening exam and follow-up studies for these 21 patients. The IQR (0.86 year) is comparable to the 36 biopsy patients. However, the median follow-up time

(2.00 years) is noticeably larger. This result can be explained from the fact that biopsy patients were screened between May of 2012 and August of 2014. At the time of publication of this paper, there simply was not time for all biopsy patients to have had two years of follow-up. Also, unlike the biopsy patients, the follow-up time for control patients was at least one year, with the minimum follow-up time being 1.02 years. The maximum follow-up time was 2.40 years.

In addition, one can understand why the control patients possessed a median follow-up time of exactly 2.00 years based on the time of the initial screening exams (May to October of 2012). If all 21 control patients were to have been screened annually, they would have returned for screening between May and October of 2013, and later, between May and October of 2014. Consequently, at the time of publication of this paper, there would have been exactly two years of follow-up for every control patient.

Figure 2(d) shows the NLF for the 21 control patients. As expected from Figure 2(b) and Figure 2(c), the NLF is higher in the 3D image than in the 2D image. At a certainty of 0.41, the NLF is 0.76 for the 2D CC view and is 1.60 for the 3D CC view, corresponding to a 109.4% increase. At the same certainty, the NLF is 0.64 for the 2D MLO view and is 1.57 for the 3D MLO view, corresponding to a 144.4% increase.

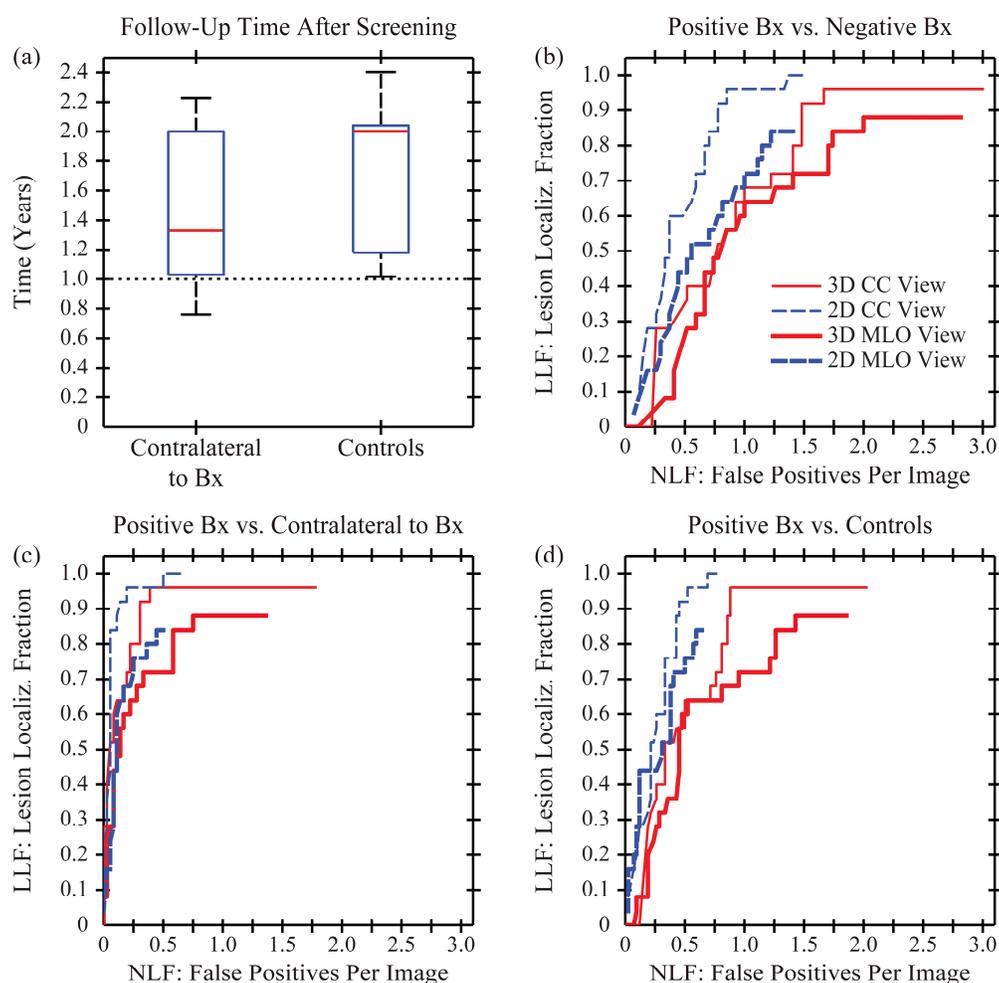


Figure 5. (a) Follow-up imaging studies can be used to determine whether cancer is detected over time. In patients with one or more follow-up studies, the longest time interval between the initial screening exam and follow-up was calculated in a boxplot. In addition, FROC curves were calculated for three different tasks; namely, distinguishing breasts possessing a positive biopsy from: (b) breasts possessing a negative biopsy, (c) breasts that are contralateral to the biopsy site, and (d) control breasts. The FROC curve for each 3D image is shifted to the right relative to the analogous curve for the 2D image. As a result, if a fixed sensitivity is specified by the user, the resultant number of FPs is larger in the 3D image than in the 2D image. Subplots (b)-(d) implicitly share a common legend.

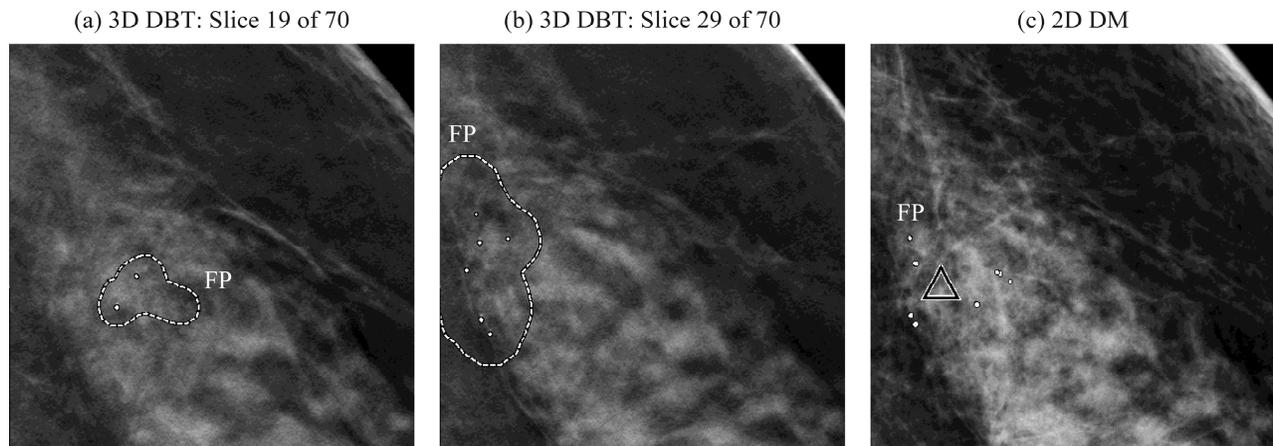


Figure 6. (a) A FP is shown in the CC view of a breast that was contralateral to a biopsy site. The FP spans slices 19 to 22. Only one slice in this range is shown for illustration. (b) An additional FP spans slices 25 to 34. (c) Because the two FPs in the 3D image project onto a similar position in the 2D image, only one FP is calculated by 2D CAD. This example illustrates why less FPs were observed in 2D CAD than in 3D CAD.

It is also demonstrated that the NLF for control breasts [Figure 2(d)] is much more similar to breasts that are contralateral to the biopsy site [Figure 2(c)] than to breasts possessing a negative biopsy [Figure 2(b)]. This result is expected since the breasts analyzed in Figure 2(c) and Figure 2(d) were not recommended for biopsy, while those in Figure 2(b) were biopsied. One can detect a subtle increase in the NLF of control breasts relative to breasts that are contralateral to the biopsy site. Analyzing why this difference exists based upon the underlying characteristics of the patients is beyond the scope of this work.

In order to understand why a larger number of FPs are detected by 3D CAD than by 2D CAD, a clinical case is shown in Figure 6. The images were obtained from a breast that was contralateral to a biopsy site. In this example, two calcification clusters are marked separately in the 3D image. The first cluster [Figure 6(a)] spans slices 19 to 22 (certainty 0.898); only one slice is shown for illustration. By contrast, the second cluster [Figure 6(b)] spans slices 25 to 34 (certainty 0.95). Because the two separate clusters in the 3D image project onto a similar position in the 2D image [Figure 6(c)], they are counted as a single mark by 2D CAD (certainty 0.893). This case illustrates how multiple clusters can superimpose onto one another in a planar projection image, yielding a lower NLF in 2D CAD than in 3D CAD.

To provide an additional explanation for the higher number of FPs in the 3D image, a case possessing a reconstruction artifact is shown in Figure 7. In this example, the breast had suspicious calcifications that were found to be benign by biopsy. A clip, which is unrelated to the biopsy site, was also present in the breast. In the 3D image, the clip is in focus in slice 41 [Figure 7(b)], and is not marked by 3D CAD. However, backprojection artifacts of the clip are marked in slices for which the clip is out of focus (certainty 0.804). While the 3D CAD mark spans slices 6 to 9, only one slice is shown for illustration [Figure 7(a)].

The positions of the backprojection artifacts can be derived by tracing rays between the clip and each x-ray source position in the DBT scan. The backprojection artifacts are oriented along the direction of x-ray tube motion for this reason. Although not shown explicitly in this paper, backprojection artifacts were also observed for clearly benign calcifications, creating the potential for additional FPs in the 3D image. One benefit of 2D DM is that there is no potential for a reconstruction artifact [Figure 7(c)]. An analysis of how the reconstruction algorithm can be perfected to minimize backprojection artifacts is beyond the scope of this work.

3.3 FROC Analysis

If the user were to select a single certainty at which to operate the CAD system, there would be an inherent pairing between the LLF and the NLF as determined by Figure 2. This pairing is shown in Figure 5(b)-(d) using free-response receiver operating characteristic (FROC) curves.¹⁰ Three different FROC curves can be created depending upon the task; namely, distinguishing breasts with a positive biopsy from: *i.* breasts with a negative biopsy [Figure 5(b)], *ii.* breasts that are contralateral to the biopsy site [Figure 5(c)], and *iii.* control breasts [Figure 5(d)].

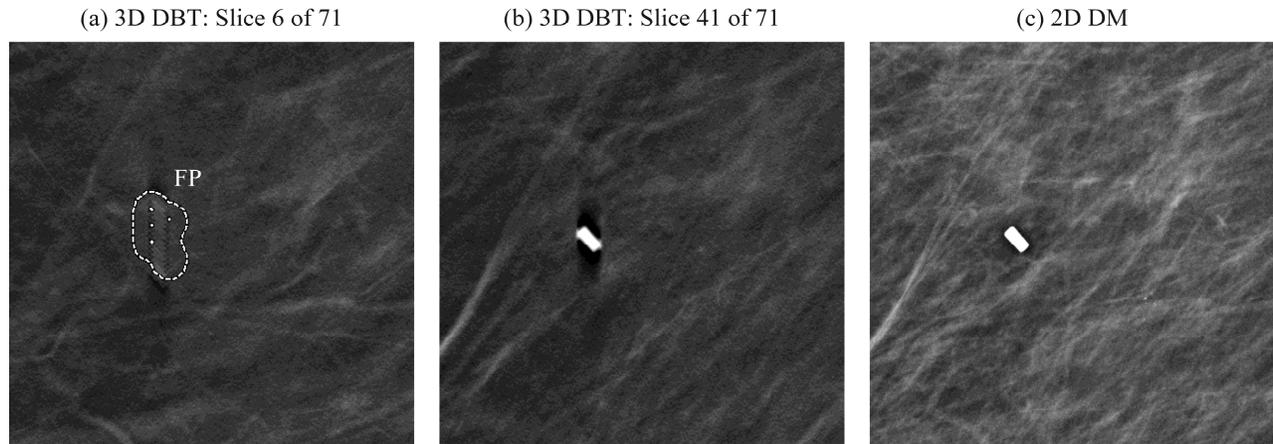


Figure 7. (a) Multiple shadows of a clip are observed in the MLO view of a breast possessing a benign biopsy. The shadows span slices 6 to 9. Only one slice in this range is shown for illustration. The positions of the shadows can be derived by tracing rays between the clip and each x-ray source position in the DBT scan. (b) The clip itself is not marked by 3D CAD in the slice for which it is in focus. (c) Because there is no potential for backprojection artifacts in 2D DM, this example illustrates why less FPs were recorded by 2D CAD than by 3D CAD.

Each FROC curve is an increasing function, indicating that the trade-off of increasing the sensitivity of the CAD system is increasing the number of FPs. To illustrate that this trade-off is more pronounced in the 3D image than in the 2D image, one can calculate the NLF at a fixed sensitivity. For the purpose of this calculation, the FROC curve that distinguishes breasts with a positive biopsy from breasts with a negative biopsy is considered [Figure 5(b)]. For a LLF of 0.80, the NLF is 0.67 for the 2D CC view and is 1.41 for the 3D CC view, corresponding to a 110.9% increase in the NLF. The analogous NLF values for the MLO view are 1.15 and 1.70, corresponding to a 48.4% increase in the NLF.

In the calculation described above, it is noteworthy that the relative increase in the NLF is much larger in the CC view than in the MLO view. Hence, the performance advantage of 2D CAD over 3D CAD is much more pronounced in the CC view. One can rationalize this finding using Figure 2(a)-(b). In the CC view, 2D CAD exhibits an advantage over 3D CAD in terms of both LLF and NLF. However, in the MLO view, 2D CAD exhibits an advantage over 3D CAD only in terms of the NLF.

In Figure 5(b)-(d), each FROC curve corresponding to a 3D image is shifted to the right relative to the FROC curve for the analogous 2D image. This result holds for all three tasks that are analyzed. As a result, the user must trade-off a larger number of FPs in the 3D image than in the 2D image, regardless of the task.

4. DISCUSSION AND CONCLUSION

This paper provides an evaluation of a prototype Hologic CAD algorithm for calcifications in DBT. Due to the newness of this 3D CAD algorithm, it is currently not FDA-approved and is intended for investigational use in the United States. The 3D CAD algorithm (version 1.0) is not as developed as the corresponding 2D CAD algorithm (version 9.4). In addition, 2D CAD includes an additional level of processing that is not yet present in 3D CAD. In particular, in 2D CAD, the marks are first calculated in each image separately, and later, the marks from multiple views of the same patient are analyzed together in an additional processing step. Despite the fact that 3D CAD is a prototype lacking this additional processing step, 3D CAD already demonstrates potential benefits over 2D CAD. In this study, the most striking benefit of 3D CAD was the modest sensitivity advantage in the MLO view [Figure 2(a)]. Also, 3D CAD allows the user to resolve the position of a calcification cluster along the direction perpendicular to the breast support. Resolution along this direction is not achievable in 2D DM due to the nature of projection imaging.

One potential drawback of 3D CAD is a larger number of FPs in comparison to 2D CAD. It should be emphasized that a 3D image has an inherently higher likelihood for FPs than a 2D image. For example, it is possible for two FPs in a 3D image to project onto a similar position in a 2D image, creating the appearance of a single FP (Figure 6). Also, it is possible for the reconstruction artifacts of a non-cancerous object, such as a clip, to be marked by 3D CAD in slices for which the object is out of focus (Figure 7). FROC curves were used to analyze the sensitivity of CAD as a function of

the mean number of FPs per image. It was shown that the user must trade-off a larger number of FPs in the 3D image than in the 2D image in order to achieve the same sensitivity. For these reasons, the most important future refinement of 3D CAD will be to minimize the number of FPs without reducing the sensitivity.

For the purpose of this work, the existence of a TP anywhere in the image was considered a lesion localization. In some cases, it was possible for a TP to mark only a small portion of the cancer as opposed to the full spatial extent of the malignancy. There was no penalty assigned to an image in which only a small portion of the cancer was marked. In future work, it will be important to analyze potential differences between 2D CAD and 3D CAD in terms of quantifying the full spatial extent of a malignancy.

One reason for the higher NLF in 3D CAD was that multiple FPs in a 3D image can project onto a similar position in a 2D image (Figure 6). Consequently, the high NLF in the 3D image can be misleading, since it is possible for multiple FPs in the 3D image to be correlated with a single FP in the 2D image. Future work will quantify the FPs in 3D CAD that do not possess a correlate in the 2D image and hence are unique to the 3D image. One such example is the clip artifact (Figure 7).

In addition to the NLF, an alternate metric that can be used to quantify the FPs is the percent of images possessing at least one FP. Unlike the NLF which penalizes each image in proportion to the number of FPs, this alternate metric assigns the same penalty to each image possessing at least one FP. Future work will analyze alternative free-response receiver operating characteristic (AFROC) curves¹⁰ resulting from the use of this metric.

In this paper, three distinct strata were identified for “negative” breasts (Figure 1). It should be emphasized that the three strata were not matched in terms of underlying characteristics which may influence the NLF; for example, density, age, and compressed breast thickness. Future work will investigate whether the lack of matching may explain differences between the strata. One such difference was the slightly higher NLF in control breasts in comparison to breasts that were contralateral to the biopsy site [Figure 2(c)-(d)].

In addition, one can divide each “negative” stratum into sub-strata based on an underlying characteristic such as density. In particular, images can be divided on the basis of high and low breast density. Future work will analyze how the NLF differs between the sub-strata. In an analogous fashion, the dependency of the LLF on an underlying characteristic such as density can also be determined. Consequently, FROC curves can be generated by calculating both the LLF and the NLF among breasts with similar underlying characteristics. This approach would allow the user to determine whether CAD performance is dependent on an underlying characteristic such as density.

On the surface, it would seem that one could pool all three “negative” strata in order to determine the NLF for all “negatives”. However, since the NLF differs between the strata, the pooled NLF would depend upon the prevalence of each stratum in the data set. For the purpose of this work, it was decided that the NLF should not be pooled in this manner, since the prevalence of each stratum in the data set is not representative of the corresponding prevalence in a screening population.

5. ACKNOWLEDGEMENT

We are grateful to David Pokrajac (Delaware State University, Dover, DE, USA) for useful discussions concerning this work. In addition, we thank Despina Kontos and Nigel Bristol (University of Pennsylvania, Philadelphia, PA, USA) for assisting with retrieving the images. Finally, we thank Ashwini Kshirsagar (Hologic, Inc., Santa Clara, CA, USA) for providing background information about ImageChecker CAD and for providing technical assistance with questions concerning this software.

Support for R.J.A. was provided by a postdoctoral fellowship grant (PDF14302589) from Susan G. Komen[®]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agency. In addition, this research project received equipment support from Hologic. E.F.C. is a lecturer for Hologic and is a member of the scientific advisory panel for Hologic. A.D.A.M. receives research support from Hologic.

This paper includes off-label uses of applications and devices not yet approved for human use in the United States. The research project was approved by the IRB at the University of Pennsylvania, and is HIPAA-compliant.

6. REFERENCES

1. Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. *Medical Physics*. 2013;40(1):014301-014301 - 014301-014312.
2. Friedewald SM, Rafferty EA, Rose SL, et al. Breast Cancer Screening Using Tomosynthesis in Combination With Digital Mammography. *Journal of the American Medical Association*. 2014;311(24):2499-2507.
3. Marshall J, Kshirsagar A, Narin S, Gkanatsios N. Will New Technologies Replace Mammography CAD as We Know It? *Lecture Notes in Computer Science*. 2014;8539:30-37.
4. 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.
5. Andersson I, Ikeda DM, Zackrisson S, et al. 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.
6. Spangler ML, Zuley ML, Sumkin JH, et al. Detection and Classification of Calcifications on Digital Breast Tomosynthesis and 2D Digital Mammography: A Comparison. *American Journal of Roentgenology*. 2011;196:320-324.
7. Kopans D, Gavenonis S, Halpern E, Moore R. Calcifications in the Breast and Digital Breast Tomosynthesis. *The Breast Journal*. 2011;17(6):638-644.
8. Svahn TM, Chakraborty DP, Ikeda D, et al. Breast tomosynthesis and digital mammography: a comparison of diagnostic accuracy. *The British Journal of Radiology*. 2012;85:e1074-e1082.
9. Rafferty EA, Park JM, Philpotts LE, et al. Assessing Radiologist Performance Using Combined Digital Mammography and Breast Tomosynthesis Compared with Digital Mammography Alone: Results of a Multicenter, Multireader Trial. *Radiology*. 2013;266(1):104-113.
10. Chakraborty DP. A Brief History of Free-Response Receiver Operating Characteristic Paradigm Data Analysis. *Academic Radiology*. 2013;20:915-919.