

Task-based Assessment of Digital Breast Tomosynthesis: Effect of Anatomy from
Multiple Anthropomorphic 3D Printed Phantoms

by

Charles Simmons Cowart

Graduate Program of Medical Physics
Duke University

Date: _____

Approved:

Joseph Y. Lo, Supervisor

Ehsan Samei

Robert Reiman

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Graduate Program of
Medical Physics in the Graduate School
of Duke University

2017

ABSTRACT

Task-based Assessment of Digital Breast Tomosynthesis: Effect of Anatomy from
Multiple Anthropomorphic 3D Printed Phantoms

by

Charles Simmons Cowart

Graduate Program of Medical Physics
Duke University

Date: _____

Approved:

Joseph Y. Lo, Supervisor

Ehsan Samei

Robert Reiman

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Graduate Program of
Medical Physics in the Graduate School of
Duke University

2017

Copyright by
Charles Simmons Cowart
2017

Abstract

Physical phantoms are an important tool in medical imaging clinical system evaluation. There exists a lack of suitable anthropomorphic physical phantoms that vary as much as a typical patient population. The lack in diversity in anthropomorphic physical phantoms makes generalizing results found using these phantoms difficult. In order to address this issue, a diverse selection of breast phantoms were 3D-printed on a Stratasys Objet350 Connex printer using tissue-approximate photopolymers. These cases were then evaluated on a clinical Hologic Selenia Dimensions Digital Breast Tomosynthesis system. The evaluation consisted of a 4-alternative-forced-choice task printed on a contrast-detail insert with silver-doped ink of concentration 200 mg/mL. The disks ranged in size from 350um-770um and a range of signal intensities was achieved by repeatedly overprinting, layering the ink 5 to 8 times. Each ink pass corresponded to an increase in signal of 1.4%. The contrast-detail insert was imaged in 8 different orientations, at a fixed kVp of 36, and varied mAs for indicated AGD, for the largest, densest case, of 1.4, 2.8, and 4.2 mGy. A channelized-Hotelling observer with Gabor channels was used for evaluation and the percent correct metric was determined. Detection performance increased as dose increased for all cases. The most dense breast case had the worst detection performance as it had the most overlapping structures to obscure the signal. The approximately average density breast and the fatty, thinner breast performed similarly, however this may be due to differences in the exposure each breast received due to differences in anatomy and the high kVp and mAs used for this

experiment. These results indicate that system performance is dependent on the anatomy being imaged. Further investigations with more phantom cases is needed to better evaluate the anatomical dependence of the system performance.

Contents

Abstract	i
List of Tables	iv
List of Figures	v
1. Introduction	1
2. Background	3
3. Methods and Materials.....	6
3.1 Breast	6
3.2 Task Assessment.....	9
3.3 Experiment	13
4. Results.....	17
5. Discussion	23
6. Conclusion	29
References	30

List of Tables

Table 1. Summary table for 3 generations of physical phantoms.	5
Table 2. System details of the Hologic Selenia Dimensions.....	14

List of Figures

Figure 1. 3 generations of 10 mm, mid-depth slabs from physical phantoms. Details summarized in table below.	4
Figure 2. Three breast cases or varied breast densities. First photo of each series is one half of the 3D printed physical phantom and the second photo of each series is the mid-depth reconstructed 1 mm slice of the physical phantom imaged using a Hologic Selenia Dimensions system.	8
Figure 3. Photographs of contrast-detail insert sandwiched in mid-depth of physical breast phantom, which is positioned and compressed like a real patient on a Hologic DBT system.	9
Figure 4. Image is a digital rendering of the 4-AFC contrast-detail insert. Dot size increases across columns and the contrast level increases across rows. Fiducials are labeled (a) – (d). Fiducials (a), (b), and (c) are used in determining the translation and rotation of the grid, and fiducial (d) is used in determining the orientation of the grid..	11
Figure 5. Image (a) is the grid imaged in uniform CIRS 50% slabs. Image (b) is the grid imaged in UM002.....	12
Figure 6. MATLAB GUI used for displaying DBT slice images, extracting relevant information from the DICOM header, detecting the location of the fiducials, and segmenting ROIs used in the model observer decision process.	16
Figure 7. Box plots of the PC taken across all signal contrast and disk size combinations indicating maximum, minimum, median and outliers. Outliers determined as those outside the 25 th and 75 th percentiles. Right-most graph shows the PC versus the Dose averaged across size and contrast. Note large differentiation at AEC and convergence to good performance at high dose.	18
Figure 8. Graphs show the change in performance as the dot size increases. The x-axis of each graph is the dot size in microns, and the y-axis of each graph is the PC. Each line in the graph represents one of the three reference techniques. Each row of graphs corresponds to a breast case, and each column of graphs represents a dot contrast level, referenced by its pass number.	19
Figure 9. Graphs show the change in performance as the dot contrast increases. The x-axis of each graph is the pass number indicating contrast level, and the y-axis of each graph is the PC. Each line in the graph represents one of the three reference techniques. Each row of graphs corresponds to a breast case, and each column of graphs represents a dot size in microns.	20

Figure 10. Graphs show direct comparison of PC between cases for all contrast levels, dot sizes, and doses. Points that fall along the dashed line indicate equal performance at that contrast and size combination, and points that fall on the side of the line indicate superior performance for the case whose side is occupied..... 22

1. Introduction

Today, breast cancer is one of the greatest risks to a woman's health with one in eight U.S. women developing breast cancer over their lifetime. In order to reduce the number of breast cancer fatalities, the Mammography Quality Standards Act (MQSA) and Program was enacted to ensure the earliest detection possible at a time when the cancer is most treatable. The screening program was set up to ensure that women in the U.S. have access to regulated, quality mammographic imaging. Early detection through screening programs and awareness is critical to catching the disease at the earliest stage.

In addition to improving awareness and education of the subject, there are many initiatives in improving breast imaging capabilities for improved detection ability and the avoidance of false-positives, uncertainties, and unnecessary callbacks. Of the improvements in breast imaging, the most important is digital breast tomosynthesis (DBT), a form of limited-angle cone-beam computed tomography based on digital mammography platforms[[1-3](#)]. Four manufacturers (Hologic, GE, Siemens, Fujifilm) have received FDA approval. Several large studies have demonstrated DBT can improve both sensitivity and specificity of breast cancer screening [[4-6](#)]. In spite of these advances, however, DBT is still defining its role in the clinic, and system quality and testing regulation are lagging and limited to vendor suggestions, which creates inconsistent performance standards. Furthermore, several other new breast imaging modalities are entering the clinic, including contrast-enhanced mammography/DBT and dedicated breast CT [[7-14](#)]. Today, there is a need for continued investigation into the

best uses for all of these new technologies, and into more advanced performance testing to better understand system capabilities and limitations.

2. Background

The use of phantoms in medical image testing is common in both clinical and research setting and directly assesses a system's performance in ways classical image quality testing is unable. For the past few decades, the most common breast imaging tests use the ACR phantom in the US and the CDMAM phantom in Europe[[15](#), [16](#)]. However, these uniform, 2D phantoms fail to mimic the complexities of breast anatomy. Researchers have added heterogeneous texture or complex layers of compartments to replicate the presence of anatomy in their phantoms [[17-20](#)], however there is still limited anthropomorphism of breast anatomy, and while presenting a more difficult challenge to readers, the task does not represent a realistic clinical challenge.

Realistic breast phantoms often begin as realistic virtual phantoms, and are primarily generated in one of, or a combination of, two methods: mathematical modeling of tissue structures [[21-23](#)] and using patient data. Our group has focused on the latter approach. First patient data from dedicated breast CT systems were segmented[[24](#)]. Following the segmentation, a large cohort of cases were made[[25](#)]. Novel physical anthropomorphic phantoms were developed using the large cohort of cases previously mentioned. These novel anthropomorphic phantoms have been developed in order to address the need for heterogeneous phantoms to better simulate and predict the clinical performance of a system.

Technological improvements in 3D printing capability and reductions in cost have awarded a unique opportunity in the development of imaging phantoms. Kiarashi

et al. first used dual-material polymer jetting to create binary phantoms with two materials representing adipose vs fibroglandular tissue[26]. Sikaria *et al.* used a commercial material mixing feature to create more realistic, “graded” blending into four tissue classes[27]. They also first introduced direct, voxelized printing to improve resolution and allow for continuous mixing of the two materials. Those previous studies were all limited by the use of an earlier cohort of breast CT data limited to resolutions inferior to imaging resolutions achievable by mammography and digital breast³. Finally in Zhao *et al.*, the latest method of voxelized 3D printing and higher resolution virtual breast phantoms based on a new cohort of breast CT data[28]. The latest generation of phantoms are approaching the resolution capabilities of breast imaging systems.

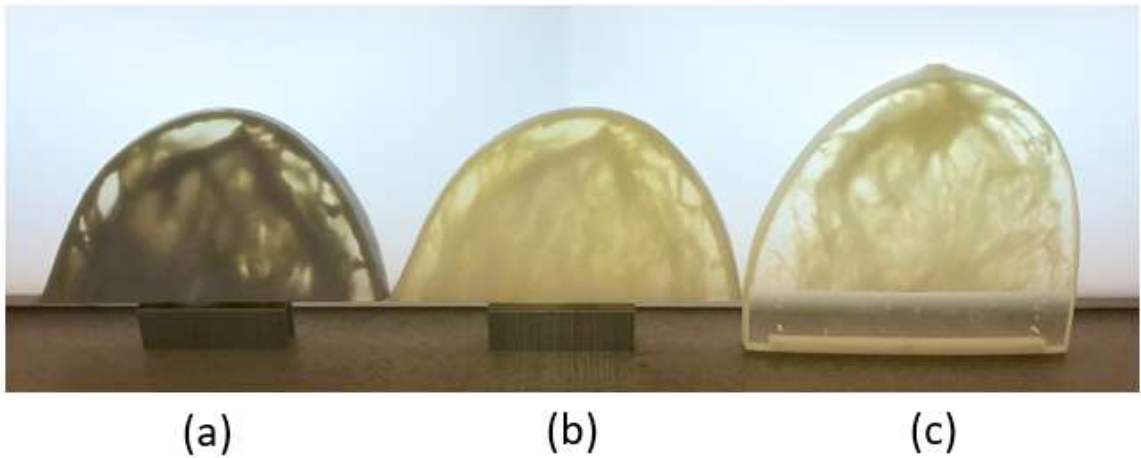


Figure 1. 3 generations of 10 mm, mid-depth slabs from physical phantoms. Details summarized in table below.

Table 1. Summary table for 3 generations of physical phantoms.

Author	Kiarashi	Sikaria	Zhao
Surfaces	Binary Surfaces	4 surfaces	voxelized & dithered
Voxel Resolution	500 μm	500 μm	155 μm

The clinical performance of DBT is commonly assessed by task-based detectability. In previous work from our group, we evaluated detectability for mammography, synthetic mammography, and DBT using a contrast-detail insert that was sandwiched in one of our anthropomorphic physical phantoms. The contrast-detail insert was created by inkjet printing with iodine-doped ink on paper. It was shown that the detection performance decreases with the introduction of complex anatomy for both human readers and observer models [29]. These studies, however, were limited to the use of a single phantom representing the anatomy of a single human subject case, which is hardly representative of the diversity of cases a radiologist will encounter.

This project expands on the previous work by evaluating detectability in DBT using higher resolution, 3D printed breast phantoms and contrast-detail inserts with higher resolution. The experiments investigate the detection capability of DBT systems with respect to dose and breast case. While we have information on how detection improves with respect to dose, we now seek to test the systems' performances as parenchymal complexity changes.

3. Methods and Materials

3.1 Breast

The breast phantoms were generated using a new cohort of patient data from a dedicated breast CT system (Koning Corporation, West Henrietta, NY) at the University of Massachusetts. Compared to the previous cohort that was limited to voxels of 500 x 500 x 500 μm , data for this study was reconstructed by filtered back projection to isotropic voxel size of 155 x 155 x 155 μm . Using our previously published denoising and segmentation algorithm, the breast CT data were converted into voxelized phantoms[25]. Multiple patients were used to establish a sample of realistic virtual phantoms with a normal distribution of breast anatomies. Based on their size, the virtual breast phantoms were finite-element compressed to thicknesses of 5 cm, 6 cm, 7 cm, and 8 cm[30]. From this population, a subset of virtual breasts phantoms with a range of thicknesses and densities were chosen to be printed. In total, three breasts were chosen from the population of cases to be printed. The breasts were chosen based on their compression thicknesses and their volumetric breast density (VBD), determined as the number of voxels in the virtual phantom labeled as glandular tissue compared to the voxel volume. The first breast, denoted at UM002, selected was 8 cm thick when compressed and had a VBD of 38.67%. The second breast, denoted at UM037, was also 8 cm when compressed, but with a VBD of 14.98%, a density similar to that of the average breast determined by Yaffe *et al.*[31]. The third, and final breast selected, denoted as UM015, was 5 cm thick and had a VBD of 10.5% to represent one of our smallest and least dense cases.

The digitally compressed virtual phantoms were printed using a dual-material photopolymer jetting printer, the Objet350 Connex (Stratasys Ltd., Rehovot, Israel), located at our institution. The photopolymers used were chosen because they provide the greatest radiographic contrast range among commercially available materials. VeroWhite was used to represent glandular tissue and TangoPlus represented adipose tissue. The two materials chosen provide the contrast range equivalent to that achieved by 36-64% breast density at mammographic energies. The phantoms were printed using the aforementioned voxelized printing method with dithered mixing of the two materials at the printer's native pitch of $42 \times 84 \times 30 \mu\text{m}$. The breasts were printed in two equal thickness slabs to improve printing speed and to allow for customization in the insertions added for task assessment or image quality measurement.

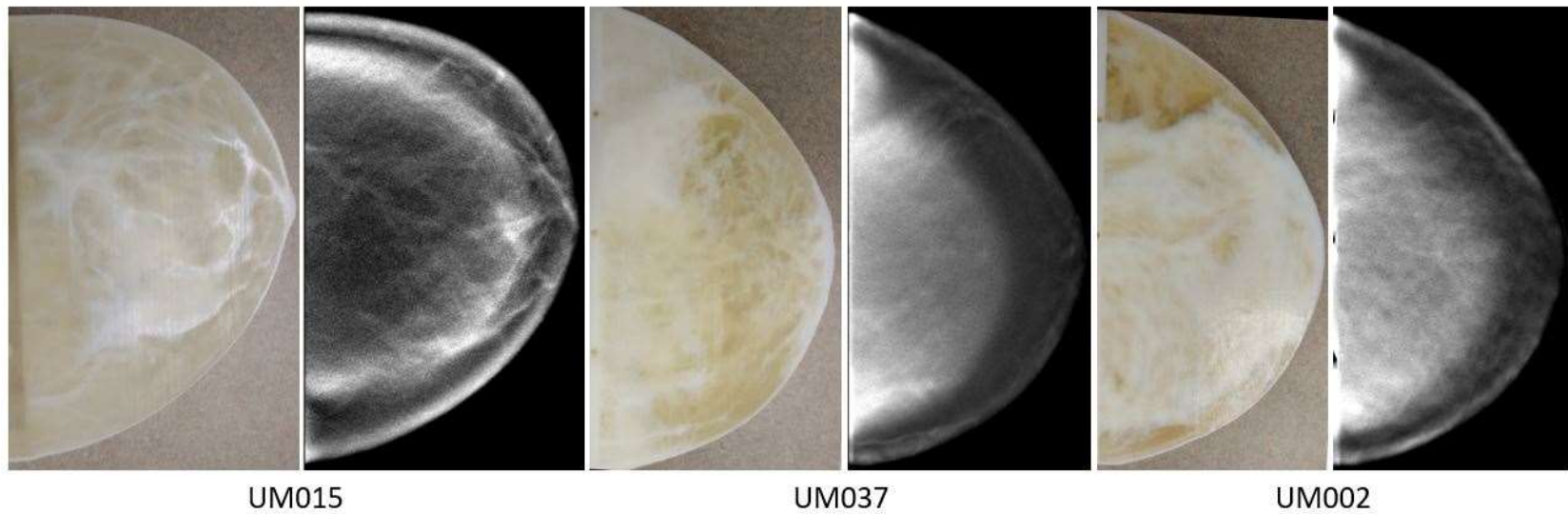


Figure 2. Three breast cases or varied breast densities. First photo of each series is one half of the 3D printed physical phantom and the second photo of each series is the mid-depth reconstructed 1 mm slice of the physical phantom imaged using a Hologic Selenia Dimensions system.

3.2 Task Assessment



Figure 3. Photographs of contrast-detail insert sandwiched in mid-depth of physical breast phantom, which is positioned and compressed like a real patient on a Hologic DBT system.

In order to assess the system performance, we selected a task based on CDMAM 4-AFC. The systems were tasked with detecting the presence of silver-doped ink dots printed on a paper insert and placed in between the two slabs of the physical breast phantoms made as described in our previous study[29]. The insert consisted of a 64 square grid of dot patterns of 350um, 490um, 630um, and 770um, and contrasts of

5.8±0.1%, 7.4±0.1%, 9±0.1%, and 10.6±0.1%, with 4 copies of each combination of size and contrast. Each square contained a center dot and one corner dot placed randomly for each square creating a 4-alternative-forced-choice (4-AFC) study.

The contrast dots were detected using a channelized Hotelling observer with Gabor channels, which was shown to have the highest correlation with human reader for this task by Ikejimba *et al.* in the same four-alternative forced-choice paradigm[29, 32]. The observer consisted of 40 channels from 2 different phases, 4 different frequencies, and 5 different angles. The edges of the passbands in cycles/pixel were 1/64, 1/32, 1/16, 1/8, and 1/4. The model observer was tasked with determining the corner which contained the randomly placed dot for each square and a percentage correct was determined from the model observer's performance for each contrast and size combination. The percent correct corresponds to an index of detectability and area under the ROC curve[33].

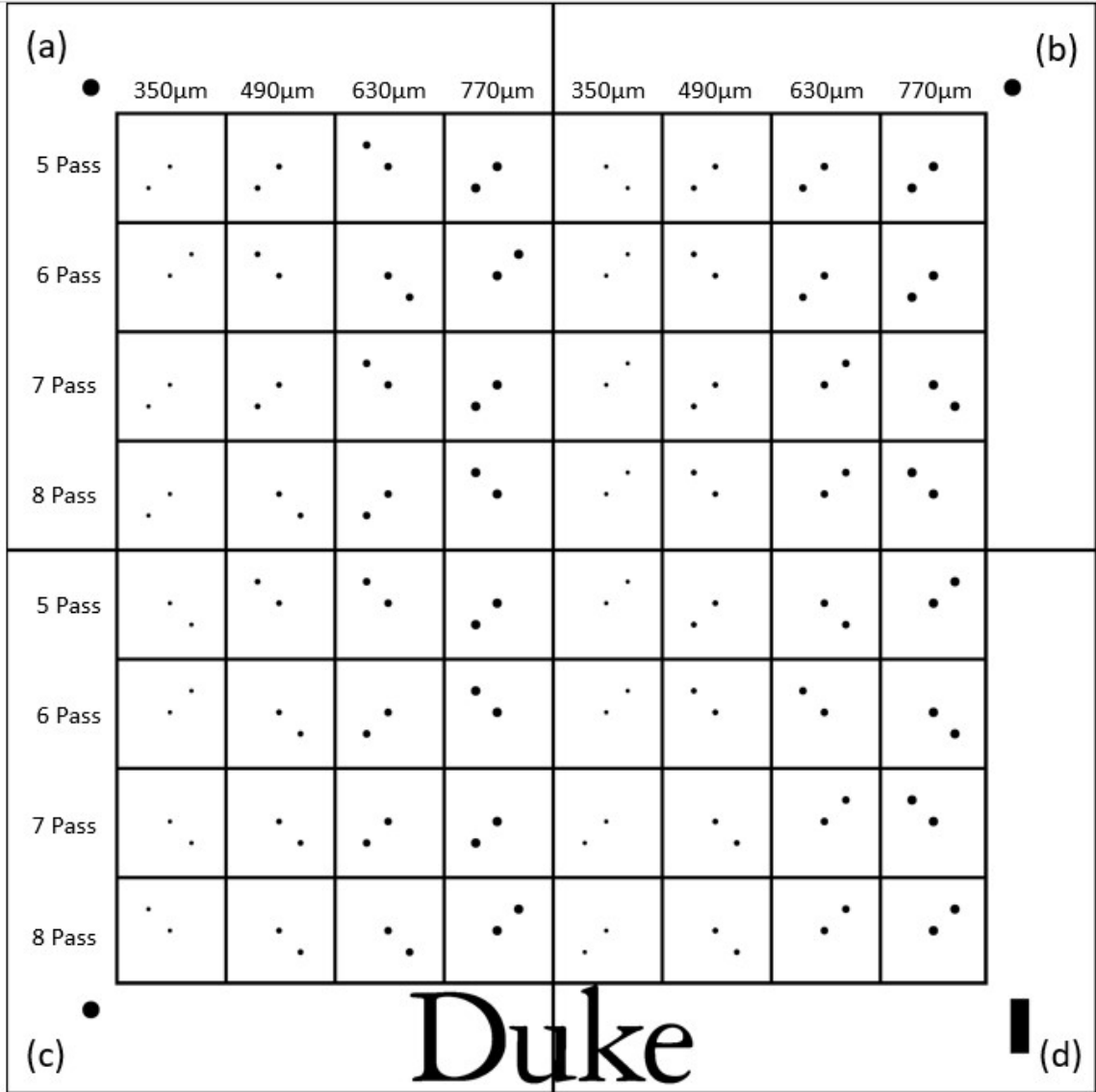


Figure 4. Image is a digital rendering of the 4-AFC contrast-detail insert. Dot size increases across columns and the contrast level increases across rows. Fiducials are labeled (a) – (d). Fiducials (a), (b), and (c) are used in determining the translation and rotation of the grid, and fiducial (d) is used in determining the orientation of the grid.

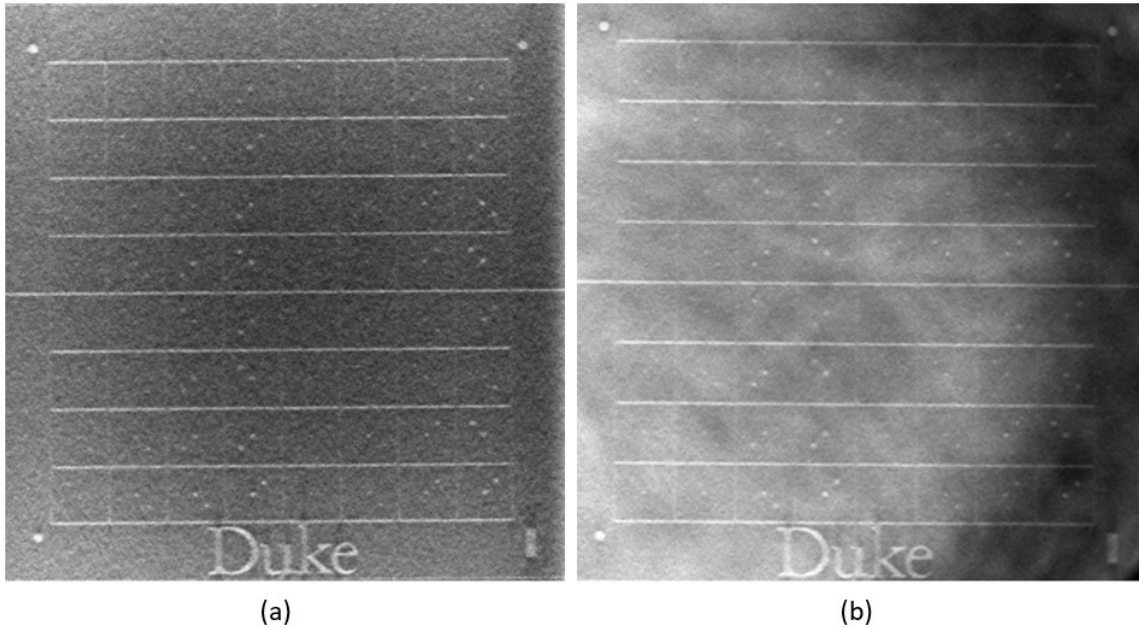


Figure 5. Image (a) is the grid imaged in uniform CIRS 50% slabs. Image (b) is the grid imaged in UM002.

The grid was printed using a commercial inkjet printer (MFC-J450DW, Brother International Corporation, Bridgewater NJ), with silver doped ink. The silver nanoparticle ink contained a silver concentration of 200 mg/mL (NBSIJ-FD02, Mitsubishi Imaging (MPM), Inc., Rye NY). Silver was used in lieu of iodine, as was previously used by Ikejimba *et al.* [29], in order to achieve high contrast at the higher kVp used in the experiment due to silver's k-edge of 25.5 keV[34]. Ranges of contrasts were achieved through repeated feeding of the contrast-detail insert through the printer. The contrast-detail insert ranged from 5 to 8 printing passes, with each pass corresponding to a 1.6% increase in signal. The insert contained high contrast fiducial markers at each corner, which were used in semi-automatic detection and segmentation of the grid.

3.3 Experiment

The breast phantoms were imaged on a clinical Hologic Selenia Dimensions at Duke University Breast Imaging Center using a Tungsten target and Aluminum filter. The tube parameters selected were based off the AEC selections for Breast 1 for a system indicated AGD of 2.8 mGy, and denoted as the reference technique, *RT*. The tube current was held constant at 36 kVp and the mAs was varied in order to achieve exposures corresponding to approximately 50%, 100%, and 150% of the reference technique. The kVp and mAs were held constant for all three breast phantoms. The kVp was fixed to maintain the same energy-dependent contrast from the contrast-detail insert. For the 5 cm thick breast phantom, the beam was filtered using 3 cm of CIRS 30% breast tissue equivalent uniform blocks attached at the tube head. This method of filtering was chosen in order to maintain the same radiographic technique and primary attenuation for this smaller phantom. The filtering technique was verified by measuring mean pixel values from several ROIs within a breast for the central projection image. The mean pixel values measured from case UM002 and UM015 differed by approximately 2% on average, indicating the breast received similar exposures. Details on the Hologic system are summarized in the table below.

Table 2. System details of the Hologic Selenia Dimensions

System	Hologic Selenia Dimension
Number of projections	15
Tube sweep (deg)	15
Source to imager distance (mm)	700
Detector element size (mm)	0.07 (2x 2 binning)
Image pixel size, x-y plane (mm) [†]	0.088
Image pixel size, z plane (mm)	1.0
Reconstruction	FBP
Target/filter	W/AI
RT Tube voltage (kV)	36
Tube load (mAs) 50%/100%/150%	45/90/138

Each breast was imaged with the insert in eight different orientations (4 rotations of 90 degrees, then turning the insert over and repeating the 4 rotations) in order to generate a greater number of test cases for the signal detection test. Each orientation presented a unique set of interactions of the targets with the surrounding anatomy. In addition to the previously mentioned orientations, each phantom was imaged with the insert absent in order to calculate image statistics used in the model observer decision process. The insert-absent images were compressed such that there was no air gap in the phantom. The probabilities of a correct decision by the model observer were compared across contrast and dot size, tissue densities, and doses in order to evaluate task-specific system performance across a variety of breast types.

The grid fiducials were detected and segmented using a graphical user interface written in MATLAB (Mathworks Inc., Natick MA). The GUI semi-automatically

detected the corner fiducials, determined the orientation and dose, segmented the ROIs for the observer model study, and provided a visual check for accuracy of the data preparation, which is crucial given that the signal targets can be as small as just a few pixels. The GUI displays the loaded tomosynthesis image and receives input from the user with regards to the case, vendor, and phantom in question. The user selects 4 positions from the image corresponding to approximate locations of the fiducials. Local areas of interest are then segmented and a combination of correlation and morphology information is used to locate the fiducials used to determine the grid orientation, translation, and rotation. Using printed fiducial markers and this GUI improved the accuracy and reproducibility of the ROI segmentation. Our previous work relied upon manually created fiducial markers and manually measured coordinates and angles, such that results were subject to high error and variability, especially at the lower doses. The previous method involved a number of empirical “shimming” values and measures to try and correct for the variability in segmentation, but these values were often subjective, tedious, and slow to determine. The current approach improved the accuracy of segmentation, provided redundant measures for checking labeling correctness, and improved the speed at which data could be prepared for use by the observer models.

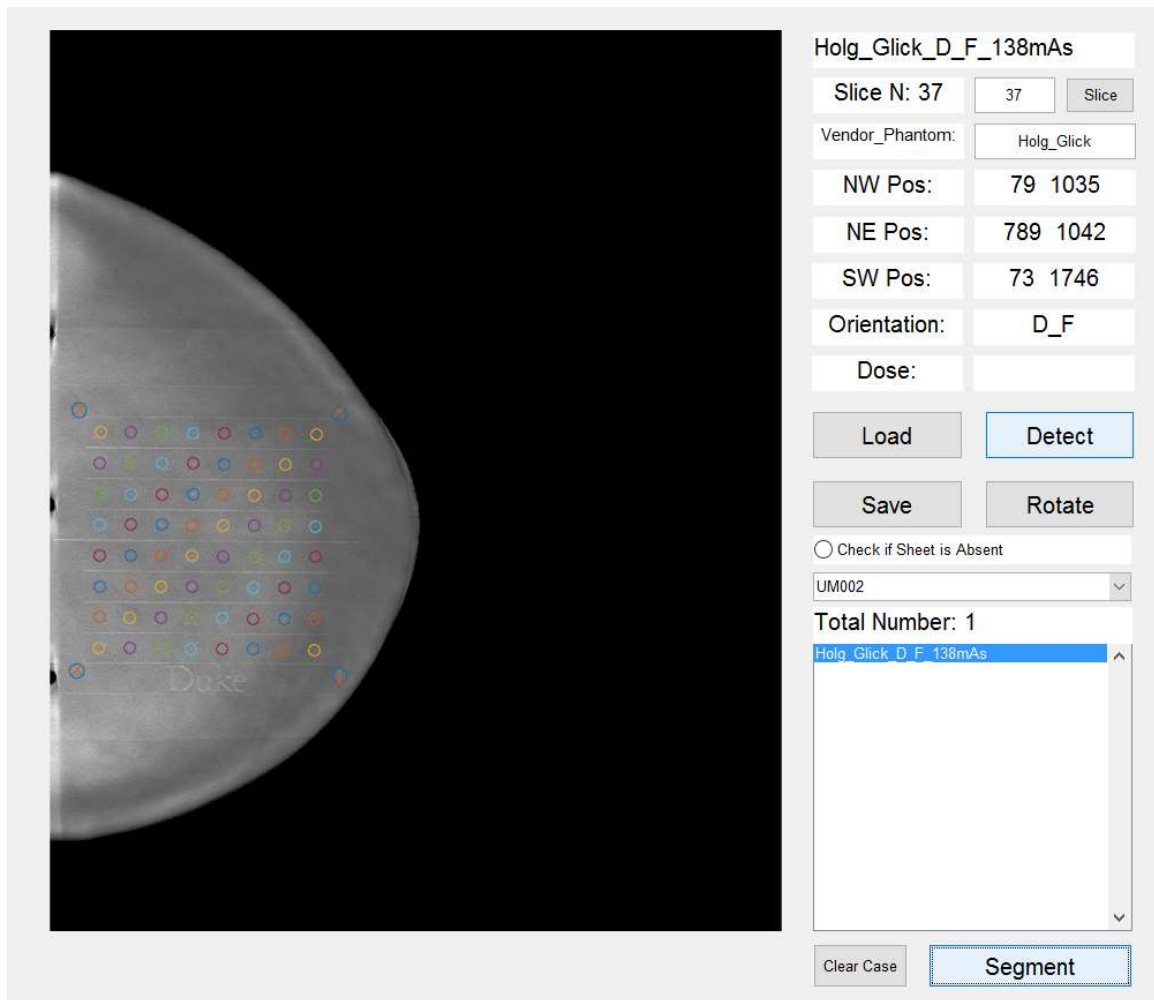


Figure 6. MATLAB GUI used for displaying DBT slice images, extracting relevant information from the DICOM header, detecting the location of the fiducials, and segmenting ROIs used in the model observer decision process.

4. Results

Observer model scores for each phantom were compared considering dose, contrast, and breast type. Results found were summarized in Figure 7. For all three cases, detection increased as dose increased. Detection performance for the densest case, UM002, was consistently lower than for the other two cases as expected due to the greater volume of glandular tissue equivalent material. The greater amount of overlapping structure in the breast makes differentiating signals difficult. The detection difference between UM015 and UM037 were small and varied across dose, signal contrast, and disk size.

Performance was found to improve with increasing size and there was no obvious relation between performance and increase in signal. There was also notable separation in performance curves between the three different references techniques. These trends are summarized in figure 8 and figure 9.

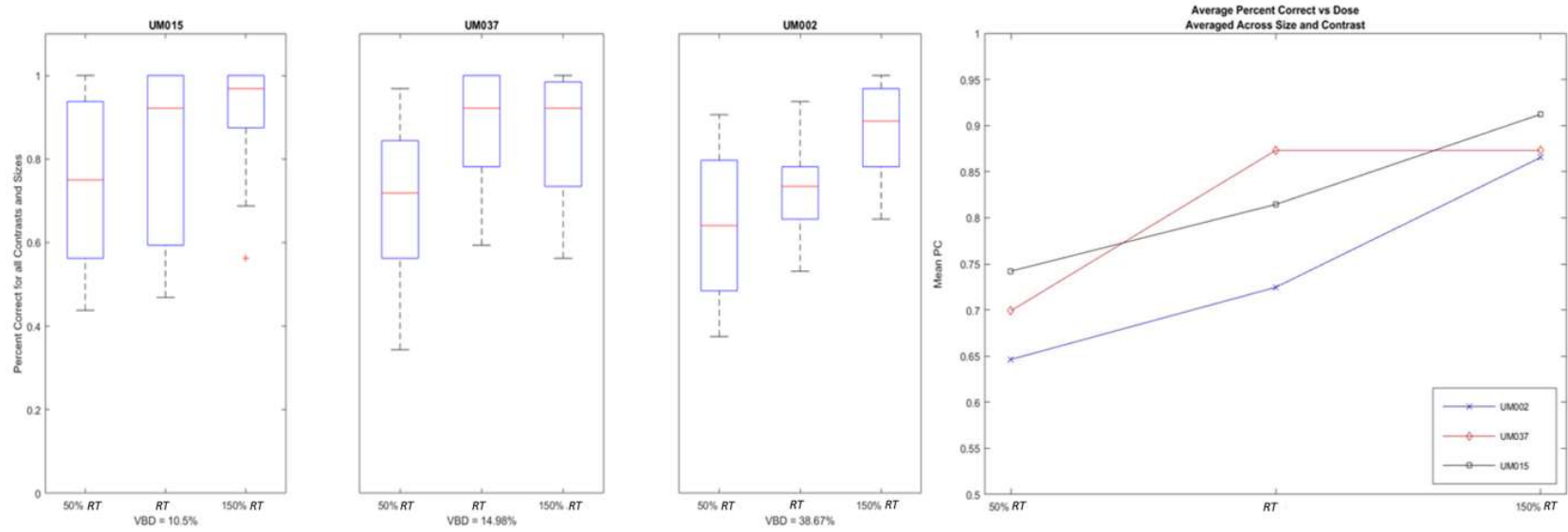


Figure 7. Box plots of the PC taken across all signal contrast and disk size combinations indicating maximum, minimum, median and outliers. Outliers determined as those outside the 25th and 75th percentiles. Right-most graph shows the PC versus the Dose averaged across size and contrast. Note large differentiation at AEC and convergence to good performance at high dose.

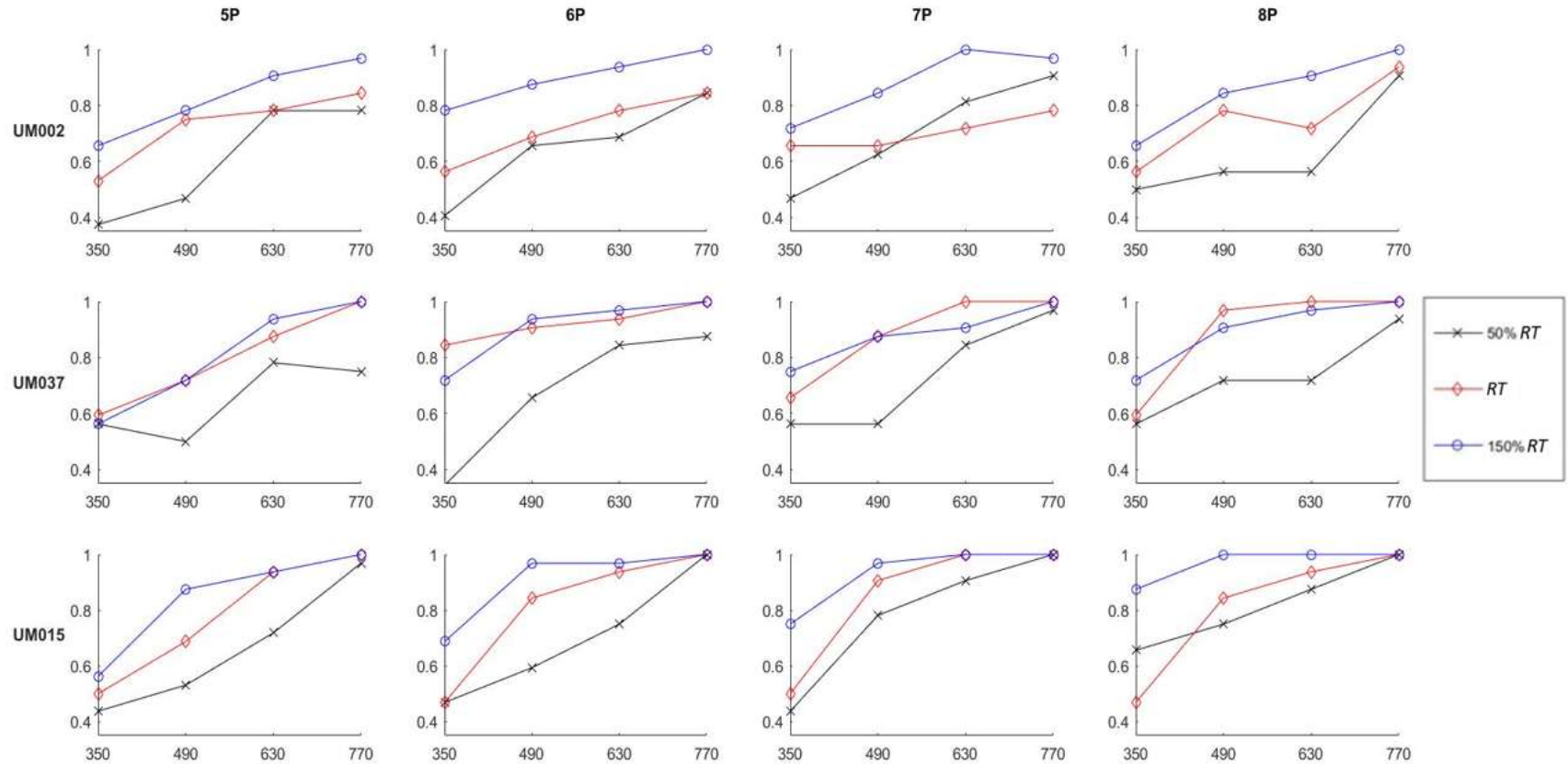


Figure 8. Graphs show the change in performance as the dot size increases. The x-axis of each graph is the dot size in microns, and the y-axis of each graph is the PC. Each line in the graph represents one of the three reference techniques. Each row of graphs corresponds to a breast case, and each column of graphs represents a dot contrast level, referenced by its pass number.

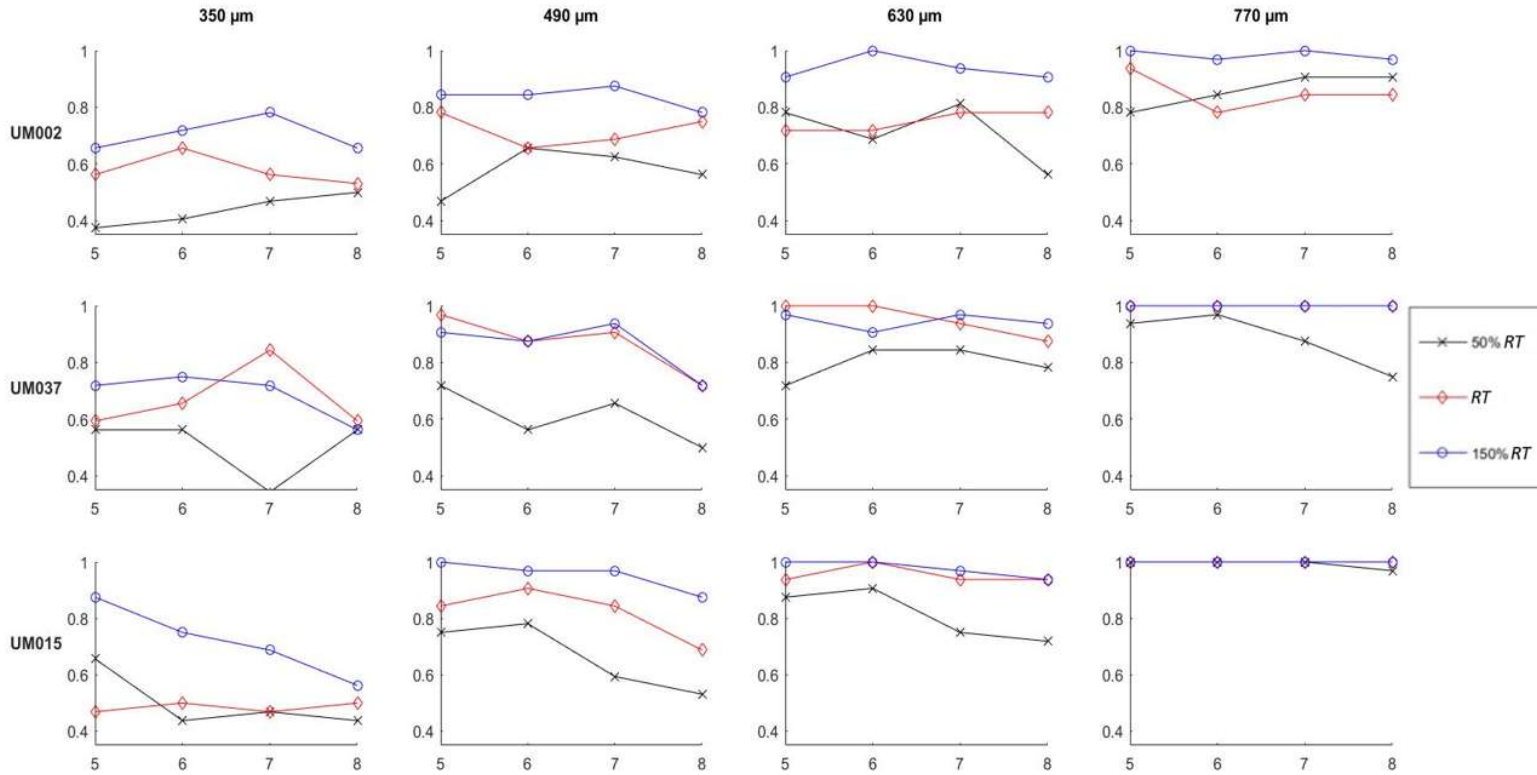


Figure 9. Graphs show the change in performance as the dot contrast increases. The x-axis of each graph is the pass number indicating contrast level, and the y-axis of each graph is the PC. Each line in the graph represents one of the three reference techniques. Each row of graphs corresponds to a breast case, and each column of graphs represents a dot size in microns.

Figure 10 further demonstrates the differences in detectability but with easier visualization of which contrast and size combinations yielded the best performance. The three graphs in figure 10 show direct comparisons between each of the cases. Corresponding size and contrast combinations PC scores are plotted with a central line bisecting the two, indicating which system performed better. UM002 had worse PC than both the other cases. UM037 and UM015 had similar PC scores for the different doses.

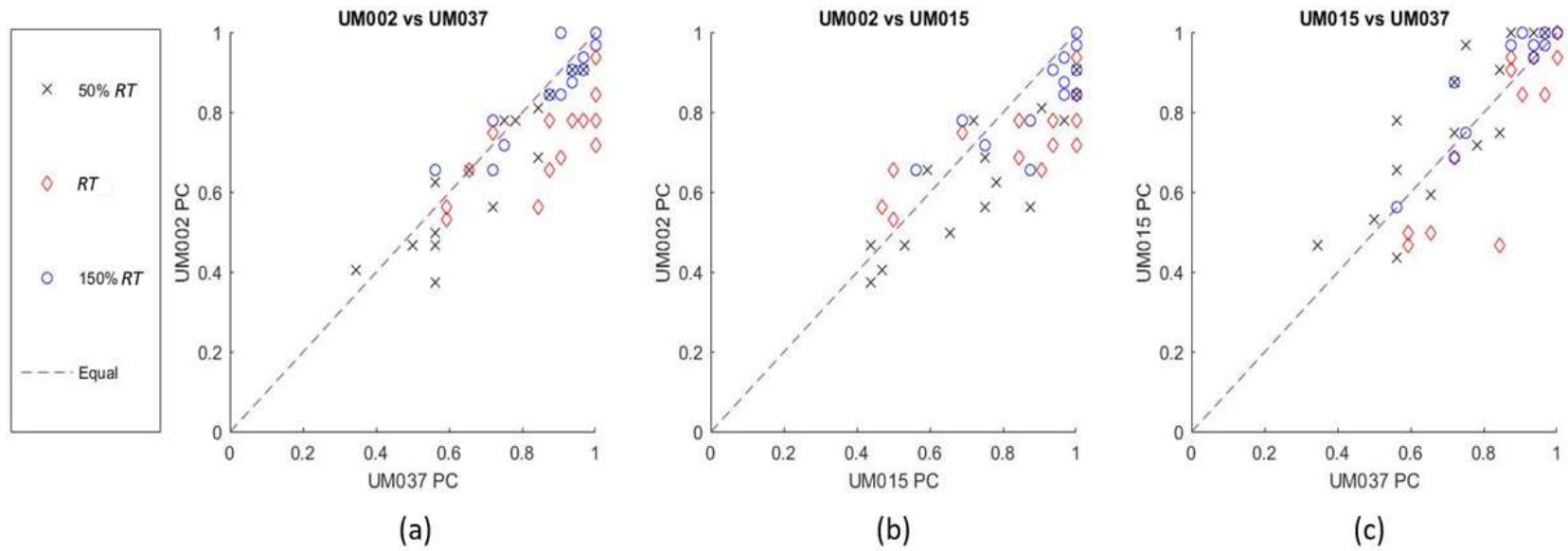


Figure 10. Graphs show direct comparison of PC between cases for all contrast levels, dot sizes, and doses. Points that fall along the dashed line indicate equal performance at that contrast and size combination, and points that fall on the side of the line indicate superior performance for the case whose side is occupied.

5. Discussion

Anthropomorphic phantoms have provided insight into system performance in ways standard quality assurance testing has been unable. Using anthropomorphic phantoms also provides the opportunity to use anatomy to study imaging systems without the ethical issues and difficulties involved with using patients.

Anthropomorphic phantoms can be imaged without regards to cumulative radiation dose and, in the case of breast imaging, without the concern of patient comfort, and still provide clinically relevant information. Previous work has shown that these phantoms can be used to measure system performance either through detectability tasks or rendition quality assessment through comparison between the virtual phantom and the image rendered by the system[[29](#), [35](#)]. Analysis using these anthropomorphic phantoms have shown quality differences between vendors and between different parameter selections, but being limited to a single phantom, the testing has been unable to show whether or not these performance differences persisted across a range of breast types. Knowledge of system performance or parameter performance for a certain breast type leads to informed clinical decisions and maximizing the quality of patient care. Using the proposed quality assessment methods and a ground truth known from the virtual phantoms, testing can be done to gain direct comparisons between systems, vendors, system performance over time, site variability, or provide insight into parameter optimization for different breast types.

This project has demonstrated the variability in performance between different breast-types for a single machine. Being able to quantify differences in detectability performance for a large, dense breast versus a large, fatty breast indicates that the dose delivered to the breast either needs to be increased to improve contrast detection, or conversely if detection is considered to be adequate then dose can be reduced. Determination of what performance is adequate and thresholds at which the objects detected aren't clinically relevant still need to be investigated before the knowledge gained can be used to inform clinical decisions. Many patient specific parameter optimizations can be decided using insights gained from the combination of clinical knowledge and of performance comparisons between different breast tissues. Using a population of anthropomorphic breast phantoms, optimization can proactively be evaluated in an exhaustive manner and inform physicians as they personalize medical treatment.

Rapid prototyping of anthropomorphic phantoms is quickly improving and soon it will be possible to generate a large population of breast phantoms for use in testing in a variety of ways. 3D printing also allows for quick and easy customization of the phantoms for the desired investigation.

In our first generation phantom, Kiarashi *et al.* used a dual-material polymer jetting printer to create an anthropomorphic phantom in a single pass [26]. The two materials were used to render surfaces into a binary phantom representing adipose vs. glandular tissue, but the absence of fine detail limited realism of the first generation of

3D printed breast phantoms. For the second generation phantom, Ikejimba *et al.* added several intermediate mixtures of the two materials, and these graded steps served to provide more realistic transitions between adipose and glandular tissues[29]. That study was also the first to introduce the inkjet-printed contrast-detail insert for task-based assessment of performance. Compared to a conventional uniform phantom, this anthropomorphic phantom resulted in lower DBT detection performance. That study indicated a need for further investigation with improved phantoms and whether the performance decrease measured was specific to the anatomy or not.

This study introduced the latest advances in the development of 3D-printed anthropomorphic breast phantoms[28]. The cases printed come from the latest cohort of breast cases which boast a much higher resolution with 155 μm voxel sizes. The printed method was adjusted to use a new voxelized printing method in order to best capitalize on the smaller virtual voxel size. The printing method also utilized a continuous dithered mixing of the two printing materials enabling a range of contrasts to be printed and better simulate actual breast parenchyma. The study also had access to an in-house printer allowing for quicker printing and troubleshooting, and ensuring quality assurance and absence of printing artifacts.

There have been a number of studies done on DBT detection performance[36-38], but these investigations have focused on the detection of micro-calcifications, which are interesting because although they have a high signal to noise ratio, but they tend to be blurred by the lower resolution of DBT. The current study seeks to evaluate detection

performance of subtle signals with mid to high frequency detail but low contrast, which may be indicative of the ability of a system to accurately render breast anatomy that is very clinically relevant, such as calcification morphology, mass margins, and distortions. The insert used in this study differed from those used previously [29] by using silver-doped ink. Silver was chosen because of its k-edge at 25.5 keV and the contrast achievable from mammographic spectra generated at the high kVp used in this study. Use of a new ink allowed for fewer printing passes required, and thus a reduction in the blurring of the dots from the printing process. Printing with a high contrast material also allowed for the accurate application of fiducials on the contrast-detail insert, unlike previously used methods involving hand-placed fiducials.

While this study improved on the previous generations of phantoms and investigated detection of subtle contrast differences, there were a number of limitations. First, there are still limits to the realism of the 3D printed breast. The contrast range achievable with currently used materials is too low, which precludes the study of very high contrast signals such as calcifications. Our group is currently working on developing new printing materials, however it is nontrivial to create new tissue-equivalent materials that are also compatible for 3D printing[28]. As the research into new materials progresses, these studies can be validated for even more realistic cases. There is also a need for further improvements in the virtual phantoms. Adding the finer details like Cooper's ligaments and ductal trees will further complicate the breast structure and better reflect a clinical image challenge[39]. There is also work being done

to model anthropomorphic lesions and inserting them into the virtual phantoms. Printing physical breast phantoms with these lesions would provide the most complete evaluation of detection performance in a clinically relevant manner[40].

These results were also limited by the small number of breast cases used. This study utilized the first three phantom prototypes that were available. Fabrication of more phantoms is underway as we continue to refine the new, voxelized printing procedure and prepare more of the new, second cohort of virtual phantom cases. While the introduction of new cases indicates performance differences are related to differences in anatomy, there needs to be a greater number of cases evaluated in order to determine what attribute has the greatest impact on system performance. A greater variety of breast phantoms also adds to its potential usefulness in studies of emerging technologies or techniques. Evaluation of performance across a large population may expedite the validation process or reduce the number of actual patients that need be exposed to radiation in the case of radiography.

Some of the results found in this project failed to demonstrate the hypothesis that larger, denser breasts would have worse detection performance than smaller, fattier breast. The similarities between the performances of UM015 and UM037 are likely a result of experimental setup. UM037 which was hypothesized to have worse detection performance than UM015 because of the greater virtual breast density and greater amount of scatter causing material, and thus greater noise for complicating the detection task. The difference between the hypothesis and the results in this study is likely due to

difference in exposures received by each breast. The smaller, fattier breast, UM015, was filtered to have a similar level exposure as the large, dense breast, UM002, as previously explained, but the larger fattier breast, UM037 had no such filtering. The reduced density in UM037 versus UM002 allowed for a higher exposure to the detector, and thus a reduction in noise and improvement in detectability performance. The filter added for UM015 didn't not add any significant scatter to the image. Differences between the large, dense breast, UM002, and the small, fatty breast, UM015, indicate that under similar exposure conditions, anatomy will affect the detection performance for this particular task assessment.

The final limitation is the evaluation of only one DBT manufacturer, when currently three others exist with FDA approval with at least one more in clinical use outside of the US (IMS) and another undergoing clinical trials (Philips). In future work, we plan to establish a comparison between multiple clinical systems in how they perform across the population of breasts. This is a time consuming process, as access to the systems requires collaboration and coordination beyond our lab and our institution.

6. Conclusion

This study was able to show that detection performance degradation is specific to anatomy, as this was the first study to use multiple cases of realistic physical breast phantoms. Detection performance was the worst for the large, high density breast as there is more overlapping structure to hide the signal tasks and more scatter from the greater amount of tissue. The observer model performed the best with the thin, fatty breast as it had fewer structures to differentiate between and less scatter. Using a population of breast cases, the study showed that performance is dependent on the individual anatomy. This work provides intriguing new avenues for investigating the patient-specific detectability in a manner that is not possible with other, existing phantoms.

References

1. Baker, J.A. and J.Y. Lo, *Breast tomosynthesis: state-of-the-art and review of the literature*. Academic radiology, 2011. **18**(10): p. 1298-1310.
2. Sechopoulos, I., *A review of breast tomosynthesis. Part I. The image acquisition process*. Medical physics, 2013. **40**(1).
3. Sechopoulos, I., *A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications*. Medical Physics, 2013. **40**(1): p. 014302-n/a.
4. Skaane, P., et al., *Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program*. Radiology, 2013. **267**(1): p. 47-56.
5. Friedewald, S.M., et al., *Breast cancer screening using tomosynthesis in combination with digital mammography*. Jama, 2014. **311**(24): p. 2499-2507.
6. Ciatto, S., et al., *Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study*. The lancet oncology, 2013. **14**(7): p. 583-589.
7. Dromain, C. and C. Balleyguier, *Contrast-Enhanced Digital Mammography*, in *Digital Mammography*, U. Bick and F. Diekmann, Editors. 2010, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 187-198.
8. Jong, R.A., et al., *Contrast-enhanced Digital Mammography: Initial Clinical Experience*. Radiology, 2003. **228**(3): p. 842-850.
9. Chen, S.C., et al., *Initial Clinical Experience With Contrast-Enhanced Digital Breast Tomosynthesis*. Academic Radiology, 2007. **14**(2): p. 229-238.
10. Lindfors, K.K., et al., *Dedicated Breast CT: Initial Clinical Experience*. Radiology, 2008. **246**(3): p. 725-733.
11. Prionas, N.D., et al., *Contrast-enhanced Dedicated Breast CT: Initial Clinical Experience*. Radiology, 2010. **256**(3): p. 714-723.
12. O'Connell, A., et al., *Cone-Beam CT for Breast Imaging: Radiation Dose, Breast Coverage, and Image Quality*. American Journal of Roentgenology, 2010. **195**(2): p. 496-509.
13. Glick, S.J., *Breast ct*. Annu. Rev. Biomed. Eng., 2007. **9**: p. 501-526.

14. McKinley, R.L., et al., *Development and Initial Demonstration of a Low-Dose Dedicated Fully 3D Breast CT System*, in *Breast Imaging: 11th International Workshop, IWDM 2012, Philadelphia, PA, USA, July 8-11, 2012. Proceedings*, A.D.A. Maidment, P.R. Bakic, and S. Gavenonis, Editors. 2012, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 442-449.
15. McLelland, R., et al., *The American College of Radiology mammography accreditation program*. AJR. American journal of roentgenology, 1991. **157**(3): p. 473-479.
16. Van Engen, R., et al., *Digital mammography update. European protocol for the quality control of the physical and technical aspects of mammography screening. S1, Part 1: acceptance and constancy testing*. European guidelines for quality assurance in breast cancer screening and diagnosis, 2013.
17. Cockmartin, L., H. Bosmans, and N. Marshall, *Comparative power law analysis of structured breast phantom and patient images in digital mammography and breast tomosynthesis*. Medical physics, 2013. **40**(8).
18. Brunner, C.C., et al. *Evaluation of various mammography phantoms for image quality assessment in digital breast tomosynthesis*. in *International Workshop on Digital Mammography*. 2012. Springer.
19. Vecchio, S., et al., *A novel approach to digital breast tomosynthesis for simultaneous acquisition of 2D and 3D images*. European radiology, 2011. **21**(6): p. 1207-1213.
20. Carton, A.K., et al., *Development of a physical 3D anthropomorphic breast phantom*. Medical physics, 2011. **38**(2): p. 891-896.
21. Bliznakova, K., et al., *Evaluation of an improved algorithm for producing realistic 3D breast software phantoms: Application for mammography*. Medical physics, 2010. **37**(11): p. 5604-5617.
22. Chen, B., et al., *An anthropomorphic breast model for breast imaging simulation and optimization*. Academic radiology, 2011. **18**(5): p. 536-546.
23. Pokrajac, D.D., A.D. Maidment, and P.R. Bakic, *Optimized generation of high resolution breast anthropomorphic software phantoms*. Medical physics, 2012. **39**(4): p. 2290-2302.
24. Li, C.M., et al., *Methodology for generating a 3D computerized breast phantom from empirical data*. Medical physics, 2009. **36**(7): p. 3122-3131.
25. Erickson, D.W., et al., *Population of 224 realistic human subject-based computational breast phantoms*. Medical physics, 2016. **43**(1): p. 23-32.

26. Kiarashi, N., et al., *Development of realistic physical breast phantoms matched to virtual breast phantoms based on human subject data*. *Medical physics*, 2015. **42**(7): p. 4116-4126.
27. Sikaria, D., et al. *Second generation anthropomorphic physical phantom for mammography and DBT: Incorporating voxelized 3D printing and inkjet printing of iodinated lesion inserts*. in *SPIE Medical Imaging*. 2016. International Society for Optics and Photonics.
28. Zhao, C., et al. *Third generation anthropomorphic physical phantom for mammography and DBT: incorporating voxelized 3D printing and uniform chest wall QC region*. 2017.
29. Ikejimba, L.C., et al., *Assessing task performance in FFDM, DBT, and synthetic mammography using uniform and anthropomorphic physical phantoms*. *Medical Physics*, 2016. **43**(10): p. 5593-5602.
30. Sturgeon, G.M., et al. *Eigenbreasts for statistical breast phantoms*. in *SPIE Medical Imaging*. 2016. International Society for Optics and Photonics.
31. Yaffe, M., et al., *The myth of the 50-50 breast*. *Medical physics*, 2009. **36**(12): p. 5437-5443.
32. Wunderlich, A., *IQmodelo: statistical software for image quality assessment with model observers*. 2014.
33. 2. *The ROC Decision Model*. *Journal of the ICRU*, 2008. **8**(1): p. 19-22.
34. Deslattes, R.D., et al., *X-ray Transition Energies (version 1.2)*. National Institute of Standards and Technology: Gaithersburg, MD, 2005.
35. Ikejimba, L., et al., *A quantitative metrology for performance characterization of five breast tomosynthesis systems based on an anthropomorphic phantom*. *Medical physics*, 2016. **43**(4): p. 1627-1638.
36. Garayoa, J., et al. *Digital breast tomosynthesis: image quality and dose saving of the synthesized image*. in *International Workshop on Digital Mammography*. 2014. Springer.
37. Gur, D., et al., *Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study*. *Academic radiology*, 2012. **19**(2): p. 166-171.
38. Chan, H.-P., et al., *Digital breast tomosynthesis: Observer performance of clustered microcalcification detection on breast phantom images acquired with an experimental*

system using variable scan angles, angular increments, and number of projection views.
Radiology, 2014. **273**(3): p. 675-685.

39. Chen, X., et al. *High-resolution, anthropomorphic, computational breast phantom: fusion of rule-based structures with patient-based anatomy.* 2017.
40. Sauer, T.J., et al. *Detectability of artificial lesions in anthropomorphic virtual breast phantoms of variable glandular fraction.* 2017.